(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: 0 148 605 A2

(12) EUROPEAN PATENT APPLICATION

(21) Application number: 84308654.7

(22) Date of filing: 12.12.84

(51) Int. Cl.⁴: C 12 N 15/00, C 12 P 21/02, C 12 N 1/00, C 12 N 5/00, C 12 Q 1/68, G 01 N 33/68, C 07 K 13/00, C 07 K 15/00, A 61 K 37/24

(30) Priority: 13.12.83 US 561024
21.02.84 US 582185
28.09.84 US 655841
30.11.84 US 675298

(43) Date of publication of application: 17.07.85 Bulletin 85/29

(84) Designated Contracting States: AT BE CH DE FR GB IT LI LU NL SE

(71) Applicant: Kirin-Amgen, Inc., 1900 Oak Terrace Lane, US-Thousand Oaks, California 91320 (US)

(72) Inventor: Lin, Fu-Kuen, 438 Thunderhead Street Thousand Oaks, California 91360 (US)

(74) Representative: Brown, John David et al, FORRESTER & BOEHMERT Widenmayerstrasse 4/I, D-8000 München 22 (DE)

(54) Production of erythropoietin.

(57) Disclosed are novel polypeptides possessing part or all of the primary structural conformation and one or more of the biological properties of mammalian erythropoietin ("EPO") which are characterized in preferred forms by being the product of procaryotic or eucaryotic host expression of an exogenous DNA sequence.

Illustratively, genomic DNA, cDNA and manufactured DNA sequences coding for part or all of the sequence of amino acid residues of EPO or for analogs thereof are incorporated into autonomously replicating plasmid or viral vectors employed to transform or transfect suitable procaryotic or eucaryotic host cells such as bacteria, yeast or vertebrate cells in culture. Upon isolation from culture media or cellular lysates or fragments, products of expression of the DNA sequences display, e.g., the immunological properties and in vitro and in vivo biological activities of EPO of human or monkey species origins. Disclosed also are chemically synthesized polypeptides sharing the biochemical and immunological properties of EPO. Also disclosed are improved methods for the detection of specific single stranded polynucleotides in a heterologous cellular or viral sample prepared from, e.g., DNA present in a plasmid or viralborne cDNA or genomic DNA "library".

EP 0 148 605 A2

"PRODUCTION OF ERYTHROPOIETIN"

This is a continuation-in-part of my co-pending U.S. Patent Application Serial Nos. 561,024, filed December 13, 1983, 582,185, filed February 21, 1984, and 655,841, filed September 28, 1984.

BACKGROUND

The present invention relates generally to the manipulation of genetic materials and, more particularly, to recombinant procedures making possible the production of polypeptides possessing part or all of the primary structural conformation and/or one or more of the biological properties of naturally-occurring erythropoietin.

A. Manipulation Of Genetic Materials

Genetic materials may be broadly defined as those chemical substances which program for and guide the manufacture of constituents of cells and viruses and direct the responses of cells and viruses. A long chain polymeric substance known as deoxyribonucleic acid (DNA) comprises the genetic material of all living cells and viruses except for certain viruses which are programmed by ribonucleic acids (RNA). The repeating units in DNA polymers are four different nucleotides, each of which consists of either a purine (adenine or guanine) or a pyrimidine (thymine or cytosine) bound to a deoxyribose sugar to which a phosphate group is attached. Attachment of nucleotides in linear polymeric form is by means of fusion of the 5' phosphate of one nucleotide to the 3' hydroxyl group of another. Functional DNA occurs in the form of stable double stranded associations of single strands of nucleotides (known as deoxyoligonucleotides),

which associations occur by means of hydrogen bonding between purine and pyrimidine bases [i.e., "complementary" associations existing either between adenine (A) and thymine (T) or guanine (G) and cytosine (C)]. By convention, nucleotides are referred to by the names of their constituent purine or pyrimidine bases, and the complementary associations of nucleotides in double stranded DNA (i.e., A-T and G-C) are referred to as "base pairs". Ribonucleic acid is a polynucleotide comprising adenine, guanine, cytosine and uracil (U), rather than thymine, bound to ribose and a phosphate group.

Most briefly put, the programming function of DNA is generally effected through a process wherein specific DNA nucleotide sequences (genes) are "transcribed" into relatively unstable messenger RNA (mRNA) polymers. The mRNA, in turn, serves as a template for the formation of structural, regulatory and catalytic proteins from amino acids. This mRNA "translation" process involves the operations of small RNA strands (tRNA) which transport and align individual amino acids along the mRNA strand to allow for formation of polypeptides in proper amino acid sequences. The mRNA "message", derived from DNA and providing the basis for the tRNA supply and orientation of any given one of the twenty amino acids for polypeptide "expression", is in the form of triplet "codons" -- sequential groupings of three nucleotide bases. In one sense, the formation of a protein is the ultimate form of "expression" of the programmed genetic message provided by the nucleotide sequence of a gene.

"Promoter" DNA sequences usually "precede" a gene in a DNA polymer and provide a site for initiation of the transcription into mRNA. "Regulator" DNA sequences, also usually "upstream" of (i.e., preceding) a gene in a given DNA polymer, bind proteins that determine the frequency (or rate) of transcriptional initiation.

Collectively referred to as "promoter/regulator" or "control" DNA sequence, these sequences which precede a selected gene (or series of genes) in a functional DNA polymer cooperate to determine whether the transcription (and eventual expression) of a gene will occur. DNA sequences which "follow" a gene in a DNA polymer and provide a signal for termination of the transcription into mRNA are referred to as transcription "terminator" sequences.

A focus of microbiological processing for the last decade has been the attempt to manufacture industrially and pharmaceutically significant substances using organisms which either do not initially have genetically coded information concerning the desired product included in their DNA, or (in the case of mammalian cells in culture) do not ordinarily express a chromosomal gene at appreciable levels. Simply put, a gene that specifies the structure of a desired polypeptide product is either isolated from a "donor" organism or chemically synthesized and then stably introduced into another organism which is preferably a self-replicating unicellular organism such as bacteria, yeast or mammalian cells in culture. Once this is done, the existing machinery for gene expression in the "transformed" or "transfected" microbial host cells operates to construct the desired product, using the exogenous DNA as a template for transcription of mRNA which is then translated into a continuous sequence of amino acid residues.

The art is rich in patent and literature publications relating to "recombinant DNA" methodologies for the isolation, synthesis, purification and amplification of genetic materials for use in the transformation of selected host organisms. U.S. Letters Patent No. 4,237,224 to Cohen, et al., for example, relates to transformation of unicellular host organisms with "hybrid" viral or circular plasmid DNA which includes

selected exogenous DNA sequences. The procedures of the Cohen, et al. patent first involve manufacture of a transformation vector by enzymatically cleaving viral or circular plasmid DNA to form linear DNA strands. Selected foreign ("exogenous" or "heterologous") DNA strands usually including sequences coding for desired product are prepared in linear form through use of similar enzymes. The linear viral or plasmid DNA is incubated with the foreign DNA in the presence of ligating enzymes capable of effecting a restoration process and "hybrid" vectors are formed which include the selected exogenous DNA segment "spliced" into the viral or circular DNA plasmid.

Transformation of compatible unicellular host organisms with the hybrid vector results in the formation of multiple copies of the exogenous DNA in the host cell population. In some instances, the desired result is simply the amplification of the foreign DNA and the "product" harvested is DNA. More frequently, the goal of transformation is the expression by the host cells of the exogenous DNA in the form of large scale synthesis of isolatable quantities of commercially significant protein or polypeptide fragments coded for by the foreign DNA. See also, e.g., U.S. Letters Patent Nos. 4,264,731 (to Shine), 4,273,875 (to Manis), 4,293,652 (to Cohen), and European Patent Application 093,619, published November 9, 1983.

The development of specific DNA sequences for splicing into DNA vectors is accomplished by a variety of techniques, depending to a great deal on the degree of "foreignness" of the "donor" to the projected host and the size of the polypeptide to be expressed in the host. At the risk of over-simplification, it can be stated that three alternative principal methods can be employed: (1) the "isolation" of double-stranded DNA sequence from the genomic DNA of the donor; (2) the chemical manufacture of

a DNA sequence providing a code for a polypeptide of interest; and (3) the <u>in vitro</u> synthesis of a double-stranded DNA sequence by enzymatic "reverse transcription" of mRNA isolated from donor cells. The last-mentioned methods which involve formation of a DNA "complement" of mRNA are generally referred to as "cDNA" methods.

Manufacture of DNA sequences is frequently the method of choice when the entire sequence of amino acid residues of the desired polypeptide product is known. DNA manufacturing procedures of co-owned, co-pending U.S. Patent Application Serial No. 483,451, by Alton, et al., (filed April 15, 1983 and corresponding to PCT US83/00605, published November 24, 1983 as WO83/04053), for example, provide a superior means for accomplishing such highly desirable results as: providing for the presence of alternate codons commonly found in genes which are highly expressed in the host organism selected for expression (e.g., providing yeast or <u>E.coli</u> "preference" codons); avoiding the presence of untranslated "intron" sequences (commonly present in mammalian genomic DNA sequences and mRNA transcripts thereof) which are not readily processed by procaryotic host cells; avoiding expression of undesired "leader" polypeptide sequences commonly coded for by genomic DNA and cDNA sequences but frequently not readily cleaved from the polypeptide of interest by bacterial or yeast host cells; providing for ready insertion of the DNA in convenient expression vectors in association with desired promoter/regulator and terminator sequences; and providing for ready construction of genes coding for polypeptide fragments and analogs of the desired polypeptides.

When the entire sequence of amino acid residues of the desired polypeptide is not known, direct manufacture of DNA sequences is not possible and isolation of DNA sequences coding for the polypeptide by a cDNA method

becomes the method of choice despite the potential drawbacks in ease of assembly of expression vectors capable of providing high levels of microbial expression referred to above. Among the standard procedures for isolating cDNA sequences of interest is the preparation of plasmid-borne cDNA "libraries" derived from reverse transcription of mRNA abundant in donor cells selected as responsible for high level expression of genes (e.g., libraries of cDNA derived from pituitary cells which express relatively large quantities of growth hormone products). Where substantial portions of the polypeptide's amino acid sequence are known, labelled, single-stranded DNA probe sequences duplicating a sequence putatively present in the "target" cDNA may be employed in DNA/DNA hybridization procedures carried out on cloned copies of the cDNA which have been denatured to single stranded form. [See, generally, the disclosure and discussions of the art provided in U.S. Patent No. 4,394,443 to Weissman, et al. and the recent demonstrations of the use of long oligonucleotide hybridization probes reported in Wallace, et al., Nuc.Acids Res., 6, pp. 3543-3557 (1979), and Reyes, et al., P.N.A.S. (U.S.A.), 79, pp. 3270-3274 (1982), and Jaye, et al., Nuc.Acids Res., 11, pp. 2325-2335 (1983). See also, U.S. Patent No. 4,358,535 to Falkow, et al., relating to DNA/DNA hybridization procedures in effecting diagnosis; published European Patent Application Nos. 0070685 and 0070687 relating to light-emitting labels on single stranded polynucleotide probes; Davis, et al., "A Manual for Genetic Engineering, Advanced Bacterial Genetics", Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1980) at pp. 55-58 and 174-176, relating to colony and plaque hybridization techniques; and, New England Nuclear (Boston, Mass.) brochures for "Gene Screen" Hybridization Transfer Membrane materials providing instruction manuals for the transfer and hybridization of DNA and RNA, Catalog No. NEF-972.]

Among the more signficant recent advances in hybridization procedures for the screening of recombinant clones is the use of labelled mixed synthetic oligonucleotide probes, each of which is potentially the complete complement of a specific DNA sequence in the hybridization sample including a heterogenous mixture of single stranded DNAs or RNAs. These procedures are acknowledged to be especially useful in the detection of cDNA clones derived from sources which provide extremely low amounts of mRNA sequences for the polypeptide of interest. Briefly put, use of stringent hybridization conditions directed toward avoidance of non-specific binding can allow, e.g., for the autoradiographic visualization of a specific cDNA clone upon the event of hybridization of the target DNA to that single probe within the mixture which is its complete complement. See generally, Wallace, et al., Nuc.Acids Res., 9, pp. 879-897 (1981); Suggs, et al. P.N.A.S. (U.S.A.), 78, pp. 6613-6617 (1981); Choo, et al., Nature, 299, pp. 178-180 (1982); Kurachi, et al., P.N.A.S. (U.S.A.), 79, pp. 6461-6464 (1982); Ohkubo, et al., P.N.A.S. (U.S.A.), 80, pp. 2196-2200 (1983); and Kornblihtt, et al. P.N.A.S. (U.S.A.), 80, pp. 3218-3222 (1983). In general, the mixed probe procedures of Wallace, et al. (1981), supra, have been expanded upon by various workers to the point where reliable results have reportedly been obtained in a cDNA clone isolation using a 32 member mixed "pool" of 16-base-long (16-mer) oligonucleotide probes of uniformly, varying DNA sequences together with a single 11-mer to effect a two-site "positive" confirmation of the presence of cDNA of interest. See, Singer-Sam, et al., P.N.A.S. (U.S.A.), 80, pp. 802-806 (1983).

The use of genomic DNA isolates is the least common of the three above-noted methods for developing

specific DNA sequences for use in recombinant procedures. This is especially true in the area of recombinant procedures directed to securing microbial expression of mammalian polypeptides and is due, principally to the complexity of mammalian genomic DNA. Thus, while reliable procedures exist for developing phage-borne libraries of genomic DNA of human and other mammalian species origins [See, e.g., Lawn, et al. Cell, 15, pp. 1157-1174 (1978) relating to procedures for generating a human genomic library commonly referred to as the "Maniatis Library"; Karn, et al., P.N.A.S. (U.S.A.), 77, pp. 5172-5176 (1980) relating to a human genomic library based on alternative restriction endonuclease fragmentation procedure; and Blattner, et al., Science, 196, pp. 161-169 (1977) describing construction of a bovine genomic library] there have been relatively few successful attempts at use of hybridization procedures in isolating genomic DNA in the absence of extensive foreknowledge of amino acid or DNA sequences. As one example, Fiddes, et al., J.Mol. and App.Genetics, 1, pp. 3-18 (1981) report the successful isolation of a gene coding for the alpha subunit of the human pituitary glycoprotein hormones from the Maniatis Library through use of a "full length" probe including a complete 621 base pair fragment of a previously-isolated cDNA sequence for the alpha subunit. As another example, Das, et al., P.N.A.S. (U.S.A.), 80, pp. 1531-1535 (1983) report isolation of human genomic clones for human HLA-DR using a 175 base pair synthetic oligonucleotide. Finally, Anderson, et al., P.N.A.S. (U.S.A.), 80, pp. 6838-6842 (1983) report the isolation of genomic clone for bovine pancreatic trypsin inhibitor (BPTI) using a single probe 86 base pairs in length and constructed according to the known amino acid sequence of BPTI. The authors note a determination of poor prospects for isolating mRNA suitable for synthesis of a cDNA library due to apparent low levels of mRNA in initially targeted parotid gland and lung tissue sources and then address the prospects of success in probing a genomic library using a mixture of labelled probes, stating: "More generally, mixed-sequence oligodeoxynucleotide probes have been used to isolate protein genes of unknown sequence from cDNA libraries. Such probes are typically mixtures of 8-32 oligonucleotides, 14-17 nucleotides in length, representing every possible codon combination for a small stretch (5-6 residues) of amino acid sequence. Under stringent hybridization conditions that discriminate against incorrectly base-paired probes, these mixtures are capable of locating specific gene sequences in clone libraries of low-to-moderate complexity. Nevertheless, because of their short length and heterogeneity, mixed probes often lack the specificity required for probing sequences as complex as a mammalian genome. This makes such a method impractical for the isolation of mammalian protein genes when the corresponding mRNAs are unavailable." (Citations omitted).

There thus continues to exist a need in the art for improved methods for effecting the rapid and efficient isolation of cDNA clones in instances where little is known of the amino acid sequence of the polypeptide coded for and where "enriched" tissue sources of mRNA are not readily available for use in constructing cDNA libraries. Such improved methods would be especially useful if they were applicable to isolating mammalian genomic clones where sparse information is available concerning amino acid sequences of the polypeptide coded for by the gene sought.

B. Erythropoietin As A Polypeptide Of Interest

Erythropoiesis, the production of red blood cells, occurs continuously throughout the human life span to offset cell destruction. Erythropoiesis is a very

precisely controlled physiological mechanism enabling sufficient numbers of red blood cells to be available in the blood for proper tissue oxygenation, but not so many that the cells would impede circulation. The formation of red blood cells occurs in the bone marrow and is under the control of the hormone, erythropoietin.

Erythropoietin, an acidic glycoprotein of approximately 34,000 dalton molecular weight, may occur in three forms: α, β and asialo. The α and β forms differ slightly in carbohydrate components, but have the same potency, biological activity and molecular weight. The asialo form is an α or β form with the terminal carbohydrate (sialic acid) removed. Erythropoietin is present in very low concentrations in plasma when the body is in a healthy state wherein tissues receive sufficient oxygenation from the existing number of erythrocytes. This normal low concentration is enough to stimulate replacement of red blood cells which are lost normally through aging.

The amount of erythropoietin in the circulation is increased under conditions of hypoxia when oxygen transport by blood cells in the circulation is reduced. Hypoxia may be caused by loss of large amounts of blood through hemorrhage, destruction of red blood cells by over-exposure to radiation, reduction in oxygen intake due to high altitudes or prolonged unconsciousness, or various forms of anemia. In response to tissues undergoing hypoxic stress, erythropoietin will increase red blood cell production by stimulating the conversion of primitive precursor cells in the bone marrow into proerythroblasts which subsequently mature, synthesize hemoglobin and are released into the circulation as red blood cells. When the number of red blood cells in circulation is greater than needed for normal tissue oxygen requirements, erythropoietin in circulation is decreased.

See generally, Testa, et al., Exp.Hematol., 8(Supp. 8), 144-152 (1980); Tong, et al., J.Biol.Chem.,

256(24), 12666-12672 (1981); Goldwasser, J.Cell.Physiol., 110(Supp. 1), 133-135 (1982); Finch, Blood, 60(6), 1241-1246 (1982); Sytowski, et al., Expt.Hematol., 8(Supp 8), 52-64 (1980: Naughton, Ann.Clin.Lab.Sci., 13(5), 432-438 (1983); Weiss, et al., Am.J.Vet.Res., 44(10),1832-1835 (1983); Lappin, et al., Exp.Hematol., 11(7), 661-666 (1983); Baciu, et al., Ann.N.Y.Acad.Sci., 414, 66-72 (1983); Murphy, et al., Acta.Haematologica Japonica, 46(7), 1380-1396 (1983); Dessypris, et al., Brit.J.Haematol., 56, 295-306 (1984); and, Emmanouel, et al., Am.J.Physiol., 247 (1 Pt 2), F168-76 (1984).

Because erythropoietin is essential in the process of red blood cell formation, the hormone has potential useful application in both the diagnosis and the treatment of blood disorders characterized by low or defective red blood cell production. See, generally, Pennathur-Das, et al., Blood, 63(5), 1168-71 (1984) and Haddy, Am.Jour.Ped.Hematol./Oncol., 4, 191-196, (1982) relating to erythropoietin in possible therapies for sickle cell disease, and Eschbach, et al. J.Clin.Invest., 74(2), pp. 434-441, (1984), describing a therapeutic regimen for uremic sheep based on *in vivo* response to erythropoietin-rich plasma infusions and proposing a dosage of 10 U EPO/kg per day for 15-40 days as corrective of anemia of the type associated with chronic renal failure. See also, Krane, Henry Ford Hosp.Med.J., 31(3), 177-181 (1983).

It has recently been estimated that the availability of erythropoietin in quantity would allow for treatment each year of anemias of 1,600,000 persons in the United States alone. See, e.g., Morrison, "Bioprocessing in Space -- an Overview", pp. 557-571 in The World Biotech Report 1984, Volume 2:USA, (Online Publications, New York, N.Y. 1984). Recent studies have provided a basis for projection of efficacy of erythro-

poietin therapy in a variety of disease states, disorders and states of hematologic irregularity: Vedovato, et al., Acta.Haematol, 71, 211-213 (1984) (beta-thalassemia); Vichinsky, et al., J.Pediatr., 105(1), 15-21 (1984) (cystic fibrosis); Cotes, et al., Brit.J.Obstet.Gyneacol., 90(4), 304-311 (1983) (pregnancy, menstrual disorders); Haga, et al., Acta.Pediatr.Scand., 72, 827-831 (1983) (early anemia of prematurity); Claus-Walker, et al., Arch.Phys.Med.Rehabil., 65, 370-374 (1984) (spinal cord injury); Dunn, et al., Eur.J.Appl.Physiol., 52, 178-182 (1984) (space flight); Miller, et al., Brit.J.Haematol., 52, 545-590 (1982) (acute blood loss); Udupa, et al., J.Lab.Clin.Med., 103(4), 574-580 and 581-588 (1984); and Lipschitz, et al., Blood, 63(3), 502-509 (1983) (aging); and Dainiak, et al., Cancer, 51(6), 1101-1106 (1983) and Schwartz, et al., Otolaryngol., 109, 269-272 (1983) (various neoplastic disease states accompanied by abnormal erythropoiesis).

Prior attempts to obtain erythropoietin in good yield from plasma or urine have proven relatively unsuccessful. Complicated and sophisticated laboratory techniques are necessary and generally result in the collection of very small amounts of impure and unstable extracts containing erythropoietin.

U.S. Letters Patent No. 3,033,753 describes a method for partially purifying erythropoietin from sheep blood plasma which provides low yields of a crude solid extract containing erythropoietin.

Initial attempts to isolate erythropoietin from urine yielded unstable, biologically inactive preparations of the hormone. U.S. Letters Patent No. 3,865,801 describes a method of stabilizing the biological activity of a crude substance containing erythropoietin recovered from urine. The resulting crude preparation containing erythropoietin purportedly retains 90% of erythropoietin activity, and is stable.

Another method of purifying human erythropoietin from urine of patients with aplastic anemia is described in Miyake, et al., J.Biol.Chem., Vol. 252, No. 15 (August 10, 1977), pp. 5558-5564. This seven-step procedure includes ion exchange chromatography, ethanol precipitation, gel filtration, and adsorption chromatography, and yields a pure erythropoietin preparation with a potency of 70,400 units/mg of protein in 21% yield.

U.S. Letters Patent No. 4,397,840 to Takezawa, et al. describes methods for preparing "an erythropoietin product" from healthy human urine specimens with weakly basic ion exchangers and proposes that the low molecular weight products obtained "have no inhibitory effects against erythropoietin.

U.K. Patent Application No. 2,085,887 by Sugimoto, et al., published May 6, 1982, describes a process for the production of hybrid human lymphoblastoid cells, reporting production levels ranging from 3 to 420 Units of erythropoietin per ml of suspension of cells (distributed into the cultures after mammalian host propagation containing up to 10^7 cells per ml. At the highest production levels asserted to have been obtained, the rate of erythropoietin production could be calculated to be from 40 to about 4,000 Units/10^6 cells/48 hours in *in vitro* culture following transfer of cells from *in vivo* propagation systems. (See also the equivalent U.S. Letters Patent No. 4,377,513.) Numerous proposals have been made for isolation of erythropoietin from tissue sources, including neoplastic cells, but the yields have been quite low. See, e.g., Jelkman, et al., Expt.Hematol., 11(7), 581-588 (1983); Tambourin, et al., P.N.A.S. (U.S.A.), 80, 6269-6273 (1983); Katsuoka, et al., Gann, 74, 534-541 (1983); Hagiwara, et al., Blood, 63(4), 828-835 (1984); and Choppin, et al., Blood, 64(2), 341-347 (1984).

Other isolation techniques utilized to obtain purified erythropoietin involve immunological procedures.

A polyclonal, serum-derived antibody directed against erythropoietin is developed by injecting an animal, preferably a rat or rabbit, with human erythropoietin. The injected human erythropoietin is recognized as a foreign antigenic substance by the immune system of the animal and elicits production of antibodies against the antigen. Differing cells responding to stimulation by the antigenic substance produce and release into circulation antibodies slightly different from those produced by other responding cells. The antibody activity remains in the serum of the animal when its blood is extracted. While unpurified serum or antibody preparations purified as a serum immunoglobulin G fraction may then be used in assays to detect and complex with human erythropoietin, the materials suffer from a major disadvantage. This serum antibody, composed of all the different antibodies produced by individual cells, is polyclonal in nature and will complex with components in crude extracts other than erythropoietin alone.

Of interest to the background of the present invention are recent advances in the art of developing continuous cultures of cells capable of producing a single species of antibody which is specifically immunologically reactive with a single antigenic determinant of a selected antigen. See, generally, Chisholm, High Technology, Vol. 3, No. 1, 57-63 (1983). Attempts have been made to employ cell fusion and hybridization techniques to develop "monoclonal" antibodies to erythropoietin and to employ these antibodies in the isolation and quantitative detection of human erythropoietin. As one example, a report of the successful development of mouse-mouse hybridoma cell lines secreting monoclonal antibodies to human erythropoietin appeared in abstract form in Lee-Huang, Abstract No. 1463 of Fed.Proc., 41, 520 (1982). As another example, a detailed description

of the preparation and use of a monoclonal, anti-erythropoietin antibody appears in Weiss, et al., P.N.A.S. (U.S.A.), 79, 5465-5469 (1982). See also, Sasaki, Biomed.Biochim.Acta., 42(11/12), S202-S206 (1983); Yanagawa, et al., Blood, 64(2), 357-364 (1984); Yanagawa, et al., J.Biol.Chem., 259(5), 2707-2710 (1984); and U.S. Letters Patent No. 4,465,624.

Also of interest to the background of the invention are reports of the immunological activity of synthetic peptides which substantially duplicate the amino acid sequence extant in naturally-occurring proteins, glycoproteins and nucleoproteins. More specifically, relatively low molecular weight polypeptides have been shown to participate in immune reactions which are similar in duration and extent to the immune reactions of physiologically significant proteins such as viral antigens, polypeptide hormones, and the like. Included among the immune reactions of such polypeptides is the provocation of the formation of specific antibodies in immunologically active animals. See, e.g., Lerner, et al., Cell, 23, 309-310 (1981); Ross, et al., Nature, 294, 654-656 (1981); Walter, et al., P.N.A.S. (U.S.A.), 77, 5197-5200 (1980); Lerner, et al., P.N.A.S. (U.S.A.), 78, 3403-3407 (1981); Walter, et al., P.N.A.S. (U.S.A.), 78, 4882-4886 (1981); Wong, et al., P.N.A.S. (U.S.A.), 78, 7412-7416 (1981); Green, et al. Cell, 28, 477-487 (1982); Nigg, et al., P.N.A.S. (U.S.A.), 79, 5322-5326 (1982); Baron, et al., Cell, 28, 395-404 (1982); Dreesman, et al., Nature, 295, 158-160 (1982); and Lerner, Scientific American, 248, No. 2, 66-74 (1983). See, also, Kaiser, et al., Science, 223, pp. 249-255 (1984) relating to biological and immunological activities of synthetic peptides which approximately share secondary structures of peptide hormones but may not share their primary structural conformation. The above studies relate, of course, to amino acid sequences of proteins other than erythro-

poietin, a substance for which no substantial amino acid sequence information has been published. In co-owned, co-pending U.S. Patent Application Serial No. 463,724, filed February 4, 1983, by J. Egrie, published August 22, 1984 as European Patent Application No. 0 116 446, there is described a mouse-mouse hybridoma cell line (A.T.C.C. No. HB8209) which produces a highly specific monoclonal, anti-erythropoietin antibody which is also specifically immunoreactive with a polypeptide comprising the following sequence of amino acids:

NH_2-Ala-Pro-Pro-Arg-Leu-Ile-Cys-Asp-Ser-Arg-Val-Leu-Glu-Arg-Tyr-Leu-Leu-Glu-Ala-Lys-COOH.

The polypeptide sequence is one assigned to the first twenty amino acid residues of mature human erythropoietin isolated according to the method of Miyake, et al., J.Biol.Chem., 252, 5558-5564 (1977) and upon which amino acid analysis was performed by the gas phase sequencer (Applied Biosystems, Inc.) according to the procedure of Hewick, M., et al., J.Biol.Chem., 256, 7990-7997 (1981). See, also, Sue, et al., Proc. Nat. Acad. Sci. (USA), 80, pp. 3651-3655 (1983) relating to development of polyclonal antibodies against a synthetic 26-mer based on a differing amino acid sequence, and Sytowski, et al., J.Immunol. Methods, 69, pp.181-186 (1984).

While polyclonal and monoclonal antibodies as described above provide highly useful materials for use in immunoassays for detection and quantification of erythropoietin and can be useful in the affinity purification of erythropoietin, it appears unlikely that these materials can readily provide for the large scale isolation of quantities of erythropoietin from mammalian sources sufficient for further analysis, clinical testing and potential wide-ranging therapeutic use of the substance in treatment of, e.g., chronic kidney disease wherein diseased tissues fail to sustain production of erythropoietin. It is consequently projected in the art that

the best prospects for fully characterizing mammalian erythropoietin and providing large quantities of it for potential diagnostic and clinical use involve successful application of recombinant procedures to effect large scale microbial synthesis of the compound.

While substantial efforts appear to have been made in attempted isolation of DNA sequences coding for human and other mammalian species erythropoietin, none appear to have been successful. This is due principally to the scarcity of tissue sources, especially human tissue sources, enriched in mRNA such as would allow for construction of a cDNA library from which a DNA sequence coding for erythropoietin might be isolated by conventional techniques. Further, so little is known of the continuous sequence of amino acid residues of erythropoietin that it is not possible to construct, e.g., long polynucleotide probes readily capable of reliable use in DNA/DNA hybridization screening of cDNA and especially genomic DNA libraries. Illustratively, the twenty amino acid sequence employed to generate the above-named monoclonal antibody produced by A.T.C.C. No. HB8209 does not admit to the construction of an unambiguous, 60 base oligonucleotide probe in the manner described by Anderson, et al., supra. It is estimated that the human gene for erythropoietin may appear as a "single copy gene" within the human genome and, in any event, the genetic material coding for human erythropoietin is likely to constitute less than 0.00005% of total human genomic DNA which would be present in a genomic library.

To date, the most successful of known reported attempts at recombinant-related methods to provide DNA sequences suitable for use in microbial expression of isolatable quantities of mammalian erythropoietin have fallen far short of the goal. As an example, Farber, et al. Exp.Hematol., 11. Supp. 14, Abstract 101 (1983) report the extraction of mRNA from kidney tissues of

phenylhydrazine-treated baboons and the injection of the mRNA into Xenopus laevis oocytes with the rather transitory result of in vitro production of a mixture of "translation products" which included among them displaying biological properties of erythropoietin. More recently, Farber, et al., Blood, 62, No. 5, Supp. No. 1, Abstract 392, at page 122a (1983) reported the in vitro translation of human kidney mRNA by frog oocytes. The resultant translation product mixture was estimated to include on the order of 220 mU of a translation product having the activity of erythropoietin per microgram of injected mRNA. While such levels of in vitro translation of exogenous mRNA coding for erythropoietin were acknowledged to be quite low (compared even to the prior reported levels of baboon mRNA translation into the sought-for product) it was held that the results confirm the human kidney as a site of erythropoietin expression, allowing for the construction of an enriched human kidney cDNA library from which the desired gene might be isolated. [See also, Farber, Clin.Res., 31(4), 769A (1983).]

Since the filing of U.S. Patent Application Serial Nos. 561,024 and 582,185, there has appeared a single report of the cloning and expression of what is asserted to have been human erythropoietin cDNA in E.coli. Briefly put, a number of cDNA clones were inserted into E.coli plasmids and β-lactamase fusion products were noted to be immunoreactive with a monoclonal antibody to an unspecified "epitope" of human erythropoietin. See, Lee-Huang, Proc. Nat. Acad. Sci. (USA), 81, pp. 2708-2712 (1984).

BRIEF SUMMARY

The present invention provides, for the first time, novel purified and isolated polypeptide products

having part or all of the primary structural conformation (i.e., continuous sequence of amino acid residues) and one or more of the biological properties (e.g., immunological properties and in vivo and in vitro biological activity) of naturally-occurring erythropoietin, including allelic variants thereof. These polypeptides are also uniquely characterized by being the product of procaryotic or eucaryotic host expression (e.g., by bacterial, yeast and mammalian cells in culture) of exogenous DNA sequences obtained by genomic or cDNA cloning or by gene synthesis. Products of microbial expression in vertebrate (e.g., mammalian and avian) cells may be further characterized by freedom from association with human proteins or other contaminants which may be associated with erythropoietin in its natural mammalian cellular environment or in extracellular fluids such as plasma or urine. The products of typical yeast (e.g., Saccaromyces cerevisiae) or procaryote (e.g., E.coli) host cells are free of association with any mammalian proteins. Depending upon the host employed, polypeptides of the invention may be glycosylated with mammalian or other eucaryotic carbohydrates or may be non-glycosylated. Polypeptides of the invention may also include an initial methionine amino acid residue (at position -1).

Novel glycoprotein products of the invention include those having a primary structural conformation sufficiently duplicative of that of a naturally-occurring (e.g., human) erythropoietin to allow possession of one or more of the biological properties thereof and having an average carbohydrate composition which differs from that of naturally-occurring (e.g., human) erythropoietin.

Vertebrate (e.g., COS-1 and CHO) cells provided by the present invention comprise the first cells ever available which can be propagated in vitro continuously and which upon growth in culture are capable of producing

in the medium of their growth in excess of 100U (preferably in excess of 500U and most preferably in excess of 1,000 to 5,000U) of erythropoietin per 10^6 cells in 48 hours as determined by radioimmunoassay.

Also provided by the present invention are synthetic polypeptides wholly or partially duplicative of continuous sequences of erythropoietin amino acid residues which are herein for the first time elucidated. These sequences, by virtue of sharing primary, secondary or tertiary structural and conformational characteristics with naturally-occurring erythropoietin may possess biological activity and/or immunological properties in common with the naturally-occurring product such that they may be employed as biologically active or immunological substitutes for erythropoietin in therapeutic and immunological processes. Correspondingly provided are monoclonal and polyclonal antibodies generated by standard means which are immunoreactive with such polypeptides and, preferably, also immunoreactive with naturally-occurring erythropoietin.

Illustrating the present invention are cloned DNA sequences of monkey and human species origins and polypeptide sequences suitably deduced therefrom which represent, respectively, the primary structural conformation of erythropoietins of monkey and human species origins.

Also provided by the present invention are novel biologically functional viral and circular plasmid DNA vectors incorporating DNA sequences of the invention and microbial (e.g., bacterial, yeast and mammalian cell) host organisms stably transformed or transfected with such vectors. Correspondingly provided by the invention are novel methods for the production of useful polypeptides comprising cultured growth of such transformed or transfected microbial hosts under conditions facilitative of large scale expression of the exogenous, vector-borne

DNA sequences and isolation of the desired polypeptides from the growth medium, cellular lysates or cellular membrane fractions.

Isolation and purification of microbially expressed polypeptides provided by the invention may be by conventional means including, e.g., preparative chromatographic separations and immunological separations involving monoclonal and/or polyclonal antibody preparations.

Having herein elucidated the sequence of amino acid residues of erythropoietin, the present invention provides for the total and/or partial manfucture of DNA sequences coding for erythropoietin and including such advantageous characteristics as incorporation of codons "preferred" for expression by selected non-mammalian hosts, provision of sites for cleavage by restriction endonuclease enzymes and provision of additional initial, terminal or intermediate DNA sequences which facilitate construction of readily expressed vectors. Correspondingly, the present invention provides for manufacture (and development by site specific mutagenesis of cDNA and genomic DNA) of DNA sequences coding for microbial expression of polypeptide analogs or derivatives of erythropoietin which differ from naturally-occurring forms in terms of the identity or location of one or more amino acid residues (i.e., deletion analogs containing less than all of the residues specified for EPO and/or substitution analogs wherein one or more residues specified are replaced by other residues and/or addition analogs wherein one or more amino acid residues is added to a terminal or medial portion of the polypeptide); and which share some or all the properties of naturally-occurring forms.

Novel DNA sequences of the invention include all sequences useful in securing expression in procaryotic or eucaryotic host cells of polypeptide products having at

least a part of the primary structural conformation and one or more of the biological properties of erythropoietin which are comprehended by: (a) the DNA sequences set out in Tables V and VI herein or their complementary strands; (b) DNA sequences which hybridize (under hybridization conditions such as illustrated herein or more stringent conditions) to DNA sequences defined in (a) or fragments thereof; and (c) DNA sequences which, but for the degeneracy of the genetic code, would hybridize to DNA sequences defined in (a) and (b) above. Specifically comprehended in part (b) are genomic DNA sequences encoding allelic variant forms of monkey and human erythropoietin and/or encoding other mammalian species of erythropoietin. Specifically comprehended by part (c) are manufactured DNA sequences encoding EPO, EPO fragments and EPO analogs which DNA sequences may incorporate codons facilitating translation of messenger RNA in non-vertebrate hosts.

Comprehended by the present invention is that class of polypeptides coded for by portions of the DNA complement to the top strand human genomic DNA sequence of Table VI herein, i.e., "complementary inverted proteins" as described by Tramontano, et al., <u>Nucleic Acids Research</u>, <u>12</u>, pp. 5049-5059 (1984).

Also comprehended by the invention are pharmaceutical compositions comprising effective amounts of polypeptide products of the invention together with suitable diluents, adjuvants and/or carriers which allow for provision of erythropoietin therapy, especially in the treatment of anemic disease states and most especially such anemic states as attend chronic renal failure.

Polypeptide products of the invention may be "labelled" by covalent association with a detectable marker substance (e.g., radiolabelled with ^{125}I) to provide reagents useful in detection and quantification of

erythropoietin in solid tissue and fluid samples such as blood or urine. DNA products of the invention may also be labelled with detectable markers (such as radiolabels and non-isotopic labels such as biotin) and employed in DNA hybridization processes to locate the erythropoietin gene position and/or the position of any related gene family in the human, monkey and other mammalian species chromosomal map. They can also be used for identifying the erythropoietin gene disorders at the DNA level and used as gene markers for identifying neighboring genes and their disorders.

As hereinafter described in detail, the present invention further provides significant improvements in methods for detection of a specific single stranded polynucleotide of unknown sequence in a heterogeneous cellular or viral sample including multiple single-stranded polynucleotides where

(a) a mixture of labelled single-stranded polynucleotide probes is prepared having uniformly varying sequences of bases, each of said probes being potentially specifically complementary to a sequence of bases which is putatively unique to the polynucleotide to be detected,

(b) the sample is fixed to a solid substrate,

(c) the substrate having the sample fixed thereto is treated to diminish further binding of polynucleotides thereto except by way of hybridization to polynucleotides in said sample,

(d) the treated substrate having the sample fixed thereto is transitorily contacted with said mixture of labelled probes under conditions facilitative of hybridization only between totally complementary polynucleotides, and,

(e) the specific polynucleotide is detected by monitoring for the presence of a hybridization reaction between it and a totally complementary probe within said

mixture of labelled probes, as evidenced by the presence of a higher density of labelled material on the substrate at the locus of the specific polynucleotide in comparison to a background density of labelled material resulting from non-specific binding of labelled probes to the substrate.

The procedures are especially effective in situations dictating use of 64, 128, 256, 512, 1024 or more mixed polynucleotide probes having a length of 17 to 20 bases in DNA/DNA or RNA/RNA or DNA/RNA hybridizations.

As described infra, the above-noted improved procedures have illustratively allowed for the identification of cDNA clones coding for erythropoietin of monkey species origins within a library prepared from anemic monkey kidney cell mRNA. More specifically, a mixture of 128 uniformly varying 20-mer probes based on amino acid sequence information derived from sequencing fractions of human erythropoietin was employed in colony hybridization procedures to identify seven "positive" erythropoietin cDNA clones within a total of 200,000 colonies. Even more remarkably, practice of the improved procedures of the invention have allowed for the rapid isolation of three positive clones from within a screening of 1,500,000 phage plaques constituting a human genomic library. This was accomplished through use of the above-noted mixture of 128 20-mer probes together with a second set of 128 17-mer probes based on amino acid analysis of a different continuous sequence of human erythropoietin.

The above-noted illustrative procedures constitute the first known instance of the use of multiple mixed oligonucleotide probes in DNA/DNA hybridization processes directed toward isolation of mammalian genomic clones and the first known instance of the use of a mixture of more than 32 oligonucleotide probes in the isolation of cDNA clones.

Numerous aspects and advantages of the invention will be apparent to those skilled in the art upon consideration of the following detailed description which provides illustrations of the practice of the invention in its presently preferred embodiments.

DETAILED DESCRIPTION

According to the present invention, DNA sequences encoding part or all of the polypeptide sequence of human and monkey species erythropoietin (hereafter, at times, "EPO") have been isolated and characterized. Further, the monkey and human origin DNA has been made the subject of eucaryotic and procaryotic expression providing isolatable quantities of polypeptides displaying biological (e.g., immunological) properties of naturally-occurring EPO as well as both *in vivo* and *in vitro* biological activities of EPO.

The DNA of monkey species origins was isolated from a cDNA library constructed with mRNA derived from kidney tissue of a monkey in a chemically induced anemic state and whose serum was immunologically determined to include high levels of EPO compared to normal monkey serum. The isolation of the desired cDNA clones containing EPO encoding DNA was accomplished through use of DNA/DNA colony hybridization employing a pool of 128 mixed, radiolabelled, 20-mer oligonucleotide probes and involved the rapid screening of 200,000 colonies. Design of the oligonucleotide probes was based on amino acid sequence information provided by enzymatic fragmentation and sequencing a small sample of human EPO.

The DNA of human species origins was isolated from a human genomic DNA library. The isolation of clones containing EPO-encoding DNA was accomplished through DNA/DNA plaque hybridization employing the above-noted pool of 128 mixed 20-mer oligonucleotide probes and

a second pool of 128 radiolabelled 17-mer probes whose sequences were based on amino acids sequence information obtained from a different enzymatic human EPO fragment.

Positive colonies and plaques were verified by means of dideoxy sequencing of clonal DNA using a subset of 16 sequences within the pool of 20-mer probes and selected clones were subjected to nucleotide sequence analysis resulting in deduction of primary structural conformation of the EPO polypeptides encoded thereby. The deduced polypeptide sequences displayed a high degree of homology to each other and to a partial sequence generated by amino acid analysis of human EPO fragments.

A selected positive monkey cDNA clone and a selected positive human genomic clone were each inserted in a "shuttle" DNA vector which was amplified in E.coli and employed to transfect mammalian cells in culture. Cultured growth of transfected host cells resulted in culture medium supernatant preparations estimated to contain as much as 3000 mU of EPO per ml of culture fluid.

The following examples are presented by way of illustration of the invention and are specifically directed to procedures carried out prior to identification of EPO encoding monkey cDNA clones and human genomic clones, to procedures resulting in such identification, and to the sequencing, development of expression systems and immunological verification of EPO expression in such systems.

More particularly, Example 1 is directed to amino acid sequencing of human EPO fragments and construction of mixtures of radiolabelled probes based on the results of this sequencing. Example 2 is generally directed to procedures involved in the identification of positive monkey cDNA clones and thus provides information concerning animal treatment and preliminary radioimmunoassay (RIA) analysis of animal sera. Example 3 is directed to the preparation of the cDNA library, colony

hybridization screening and verification of positive clones, DNA sequencing of a positive cDNA clone and the generation of monkey EPO polypeptide primary structural conformation (amino acid sequence) information. Example 4 is directed to procedures involved in the identification of positive human genomic clones and thus provides information concerning the source of the genomic library, plaque hybridization procedures and verification of positive clones. Example 5 is directed to DNA sequencing of a positive genomic clone and the generation of human EPO polypeptide amino acid sequence information including a comparison thereof to the monkey EPO sequence information. Example 6 is directed to procedures for construction of a vector incorporating EPO-encoding DNA derived from a positive monkey cDNA clone, the use of the vector for transfection of COS-1 cells and cultured growth of the transfected cells. Example 7 is directed to procedures for construction of a vector incorporating EPO-encoding DNA derived from a positive human genomic clone, the use of the vector for transfection of COS-1 cells and the cultured growth of the transfected cells. Example 8 is directed to immunoassay procedures performed on media supernatants obtained from the cultured growth of transfected cells according to Example 6 and 7. Example 9 is directed to in vitro and in vivo biological activity of microbially expressed EPO of Examples 6 and 7.

Example 10 is directed to a development of mammalian host expression systems for monkey species EPO cDNA and human species genomic DNA involving Chinese hamster ovary ("CHO") cells and to the immunological and biological activities of products of these expression systems as well as characterization of such products. Example 11 is directed to the preparation of manufactured genes encoding human species EPO and EPO analogs, which genes include a number of preference codons for

expression in <u>E.coli</u> and yeast host cells, and to expression systems based thereon. Example 12 relates to the immunological and biological activity profiles of expression products of the systems of Example 11.

EXAMPLE 1

A. Human EPO Fragment Amino Acid Sequencing

Human EPO was isolated from urine and subjected to tryptic digestion resulting in the development and isolation of 17 discrete fragments in quantities approximating 100-150 picomoles.

Fragments were arbitrarily assigned numbers and were analyzed for amino acid sequence by microsequence analysis using a gas phase sequencer (Applied Biosystems) to provide the sequence information set out in Table I, below, wherein single letter codes are employed and "X" designates a residue which was not unambiguously determined.

TABLE I

Fragment No.	Sequence Analysis Result
T4a	A-P-P-R
T4b	G-K-L-K
T9	A-L-G-A-Q-K
T13	V-L-E-R
T16	A-V-S-G-L-R
T18	L-F-R
T21	K-L-F-R
T25	Y-L-L-E-A-K
T26a	L-I-C-D-S-R
T26b	L-Y-T-G-E-A-C-R
T27	T-I-T-A-D-T-F-R
T28	E-A-I-S-P-P-D-A-A-M-A-A-P-L-R
T30	E-A-E-X-I-T-T-G-X-A-E-H-X-S-L-N-E-X-I-T-V-P
T31	V-Y-S-N-F-L-R
T33	S-L-T-T-L-L-R
T35	V-N-F-Y-A-W-K
T38	G-Q-A-L-L-V-X-S-S-Q-P-W-E-P-L-Q-L-H-V-D-K

B. Design and Construction of Oligonucleotide Probe Mixtures

The amino acid sequences set out in Table I were reviewed in the context of the degeneracy of the genetic code for the purpose of ascertaining whether mixed probe procedures could be applied to DNA/DNA hybridization procedures on cDNA and/or genomic DNA libraries. This analysis revealed that within Fragment No. T35 there existed a series of 7 amino acid residues (Val-Asn-Phe-Tyr-Ala-Trp-Lys) which could be uniquely characterized as encoded for by one of 128 possible DNA sequences spanning 20 base pairs. A first set of 128 20-mer oligonucleotides was therefore synthesized by standard phosphoamidite methods (See, e.g., Beaucage, et al., Tetrahedron Letters, 22, pp. 1859-1862 (1981) on a solid support according to the sequence set out in Table II, below.

TABLE II

Residue -	Val -	Asn	Phe	Tyr	Ala	Trp	Lys	
3'	CAA	TTG	AAG	ATG	CGA	ACC	TT	- 5'
	T	A	A	A	T			
	G				G			
	C				C			

Further analysis revealed that within fragment No. T38 there existed a series of 6 amino acid residues (Gln-Pro-Trp-Glu-Pro-Leu) on the basis of which there could be prepared a pool of 128 mixed olignucleotide 17-mer probes as set out in Table III, below.

TABLE III

Residue -	Gln	Pro	Trp	Glu	Pro	Leu	
3'	GTT	GGA	ACC	CTT	GGA	GA	- 5'
	C	T		C	T	A	
		G			G		
		C			C		

Oligonucleotide probes were labelled at the 5' end with gamma - ^{32}P-ATP, 7500-8000 Ci/mmole (ICN) using T_4 polynucleotide kinase (NEN).

EXAMPLE 2

A. Monkey Treatment Procedures and RIA Analysis

Female Cynomolgus monkeys *Macaca fascicularias* (2.5-3 kg, 1.5-2 years old) were treated subcutaneously with a pH 7.0 solution of phenylhydrazine hydrochloride at a dosage level of 12.5 mg/kg on days 1, 3 and 5. The hematocrit was monitored prior to each injection. On day 7, or whenever the hematocrit level fell below 25% of the initial level, serum and kidneys were harvested after administration of 25 mg/kg doses of ketamine hydrochloride. Harvested materials were immediately frozen in liquid nitrogen and stored at -70°C.

B. RIA for EPO

Radioimmunoassay procedures applied for quantitative detection of EPO in samples were conducted according to the following procedures:

An erythropoietin standard or unknown sample was incubated together with antiserum for two hours at 37°C. After the two hour incubation, the sample tubes were cooled on ice, ^{125}I-labelled erythropoietin was added, and the tubes were incubated at 0°C for at least 15 more hours. Each assay tube contained 500 µl of incubation mixture consisting of 50 µl of diluted immune sera, 10,000 cpm of ^{125}I-erythropoietin, 5 µl trasylol and 0-250 µl of either EPO standard or unknown sample, with PBS containing 0.1% BSA making up the remaining volume. The antiserum used was the second test bleed of a rabbit

immunized with a 1% pure preparation of human urinary erythropoietin. The final antiserum dilution on the assay was adjusted so that the antibody-bound ^{125}I-EPO did not exceed 10-20% of the input total counts. In general, this corresponded to a final antiserum dilution of from 1:50,000 to 1:100,000.

The antibody-bound ^{125}I-erythropoietin was precipitated by the addition of 150 µl Staph A. After a 40 min. incubation, the samples were centrifuged and the pellets were washed two times with 0.75 ml 10 mM Tris-HCl pH 8.2 containing 0.15M NaCl, 2mM EDTA, and 0.05% Triton X-100. The washed pellets were counted in a gamma counter to determine the percent of ^{125}I-erythropoietin bound. Counts bound by pre-immune sera were subtracted from all final values to correct for nonspecific precipitation. The erythropoietin content of the unknown samples was determined by comparison to the standard curve.

The above procedure was applied to monkey serum obtained in Part A, above, as well as to the untreated monkey serum. Normal serum levels were assayed to contain approximately 36 mU/ml while treated monkey serum contained from 1000 to 1700 mU/ml.

EXAMPLE 3

A. Monkey cDNA Library Construction

Messenger RNA was isolated from normal and anemic monkey kidneys by the guanidinium thiocyanate procedure of Chirgwin, et al., Biochemistry, 18, p. 5294 (1979) and poly $(A)^+$ mRNA was purified by two runs of oligo(dT)-cellulose column chromatography as described at pp. 197-198 in Maniatis, et al., "Molecular Cloning, A Laboratory Manual" (Cold Springs Harbor Laboratory, Cold Springs, Harbor, N.Y., 1982). The cDNA library was constructed according to a modification of the general pro-

cedures of Okayama, et al., Mol. and Cell.Biol., 2, pp. 161-170 (1982). The key features of the presently preferred procedures were as follows: (1) pUC8 was used as the sole vector, cut with PstI and then tailed with oligo dT of 60-80 bases in length; (2) HincII digestion was used to remove the oligo dT tail from one end of the vector; (3) first strand synthesis and oligo dG tailing was carried out according to the published procedure; (4) BamHI digestion was employed to remove the oligo dG tail from one end of the vector; and (5) replacement of the RNA strand by DNA was in the presence of two linkers (GATCTAAAGACCGTCCCCCCCCC and ACGGTCTTTA) in a three-fold molar excess over the oligo dG tailed vector.

B. Colony Hybridization Procedures For Screening Monkey cDNA Library

Transformed E.coli were spread out at a density of 9000 colonies per 10 x 10 cm plate on nutrient plates containing 50 micrograms/ml Ampicillin. GeneScreen filters (New England Nuclear Catalog No. NEF-972) were pre-wet on a BHI-CAM plate (Bacto brain heart infusion 37 g/L, Casamino acids 2 g/L and agar 15 g/L, containing 500 micrograms/ml Chloramphenicol) and were used to lift the colonies off the plate. The colonies were grown in the same medium for 12 hours or longer to amplify the plasmid copy numbers. The amplified colonies (colony side up) were treated by serially placing the filters over 2 pieces of Whatman 3 MM paper saturated with each of the following solutions:

(1) 50 mM glucose - 25 mM Tris-HCl (pH 8.0) - 10 mM EDTA (pH 8.0) for five minutes;

(2) 0.5 M NaOH for ten minutes; and

(3) 1.0 M Tris-HCl (pH 7.5) for three minutes.

The filters were then air dried in a vacuum over at 80°C for two hours.

The filters were then subjected to Proteinase K

digestion through treatment with a solution containing 50 micrograms/ml of the protease enzyme in Buffer K [0.1M Tris-HCl (pH 8.0) - 0.15M NaCl - 10 mM EDTA (pH 8.2) -0.2% SDS]. Specifically, 5 ml of the solution was added to each filter and the digestion was allowed to proceed at 55°C for 30 minutes, after which the solution was removed.

The filters were then treated with 4 ml of a prehybridization buffer (5 x SSPE - 0.5% SDS - 100 micrograms/ml SS E.coli DNA - 5 x BFP). The prehybridization treatment was carried out at 55°C, generally for 4 hours or longer, after which the prehybridization buffer was removed.

The hybridization process was carried out in the following manner. To each filter was added 3 ml of hybridization buffer (5 x SSPE - 0.5% SDS - 100 micrograms/ml yeast tRNA) containing 0.025 picomoles of each of the 128 probe sequences of Table II (the total mixture being designated the EPV mixture) and the filters were maintained at 48°C for 20 hours. This temperature was 2°C less than the lowest of the calculated dissociation temperatures (Td) determined for any of the probes.

Following hybridization, the filters were washed three times for ten minutes on a shaker with 6 x SSC -0.1% SDS at room temperature and washed two to three times with 6 x SSC - 1% SDS at the hybridization temperature (48°C).

Autoradiography of the filters revealed seven positive clones among the 200,000 colonies screened.

Initial sequence analysis of one of the putative monkey cDNA clones (designated clone 83) was performed for verification purposes by a modification of the procedure of Wallace, et al., Gene, 16, pp. 21-26 (1981). Briefly, plasmid DNA from monkey cDNA clone 83 was linearized by digestion with EcoRI and denatured by

heating in a boiling water bath. The nucleotide sequence was determined by the dideoxy method of Sanger, et al., P.N.A.S. (U.S.A.), 74, pp. 5463-5467 (1977). A subset of the EPV mixture of probes consisting of 16 sequences was used as a primer for the sequencing reactions.

C. Monkey EPO cDNA Sequencing

Nucleotide sequence analysis of clone 83 was carried out by the procedures of Messing, *Methods in Enzymology*, 101, pp. 20-78 (1983). Set out in Table IV is a preliminary restriction map analysis of the approximately 1600 base pair EcoRI/HindIII cloned fragment of clone 83. Approximate locations of restriction endonuclease enzyme recognition sites are provided in terms of number of bases 3' to the EcoRI site at the 5' end of the fragment. Nucleotide sequencing was carried out by sequencing individual restriction fragments with the intent of matching overlapping fragments. For example, an overlap of sequence information provided by analysis of nucleotides in a restriction fragment designated C113 (Sau3A at ~111/SmaI at ~324) and the reverse order sequencing of a fragment designated C73 (AluI at ~424/BstEII at ~203).

TABLE IV

Restriction Enzyme Recognition Site	Approximate Location(s)
EcoRI	1
Sau3A	111
SmaI	180
BstEII	203
SmaI	324
KpnI	371
RsaI	372
AluI	424
PstI	426
AluI	430
HpaI	466
AluI	546
PstI	601
PvuII	604
AluI	605
AluI	782
AluI	788
RsaI	792
PstI	807
AluI	841
AluI	927
NcoI	946
Sau3A	1014
AluI	1072
AluI	1115
AluI	1223
PstI	1301
RsaI	1343
AluI	1384
HindIII	1449
AluI	1450
HindIII	1585

Sequencing of approximately 1342 base pairs (within the region spanning the Sau3A site 3' to the EcoRI site and the HindIII site) and analysis of all possible reading frames has allowed for the development of DNA and amino acid sequence information set out in Table V. In the Table, the putative initial amino acid residue of the amino terminal of mature EPO (as verified by correlation to the previously mentioned sequence analysis of twenty amino terminal residues) is designated by the numeral +1. The presence of a methionine-specifying ATG codon (designated -27) "upstream" of the initial amino terminal alanine residue as the first residue designated for the amino acid sequence of the mature protein is indicative of the likelihood that EPO is initially expressed in the cytoplasm in a precursor form including a 27 amino acid "leader" region which is excised prior to entry of mature EPO into circulation. Potential glycosylation sites within the polypeptide are designated by asterisks. The estimated molecular weight of the translated region was determine to be 21,117 daltons and the M.W. of the 165 residues of the polypeptide constituting mature monkey EPO was determined to be 18,236 daltons.

TABLE V

Translation of Monkey EPO cDNA

```
Sau3A
GATCCCGCGCCCCCTGGACAGCCGCCCTCTCCTCCAGGCCCGTGGGGCTGGCCCTGCCC

CGCTGAACTTCCCGGGATGAGGACTCCCGGTGTGGTCACCGCGCGCCTAGGTCGCTGAG

                           -27                                 -20
                           Met Gly Val His Glu Cys Pro Ala Trp
GGACCCCGGCCAGGCGCGGAGATG GGG GTG CAC GAA TGT CCT GCC TGG

                                    -10
Leu Trp Leu Leu Leu Ser Leu Val Ser Leu Pro Leu Gly Leu Pro
CTG TGG CTT CTC CTG TCT CTC GTG TCG CTC CCT CTG GGC CTC CCA

        -1  +1                                  10
Val Pro Gly Ala Pro Pro Arg Leu Ile Cys Asp Ser Arg Val Leu
GTC CCG GGC GCC CCA CCA CGC CTC ATC TGT GAC AGC CGA GTC CTG

                            20              *
Glu Arg Tyr Leu Leu Glu Ala Lys Glu Ala Glu Asn Val Thr Met
GAG AGG TAC CTC TTG GAG GCC AAG GAG GCC GAG AAT GTC ACG ATG

        30                              *       40
Gly Cys Ser Glu Ser Cys Ser Leu Asn Glu Asn Ile Thr Val Pro
GGC TGT TCC GAA AGC TGC AGC TTG AAT GAG AAT ATC ACC GTC CCA
```


TABLE V (continuted)

```
                            50
Asp Thr Lys Val Asn Phe Tyr Ala Trp Lys Arg Met Glu Val Gly
GAC ACC AAA GTT AAC TTC TAT GCC TGG AAG AGG ATG GAG GTC GGG

        60                                  70
Gln Gln Ala Val Glu Val Trp Gln Gly Leu Ala Leu Leu Ser Glu
CAG CAG GCT GTA GAA GTC TGG CAG GGC CTG GCC CTG CTC TCA GAA

                            80          *
Ala Val Leu Arg Gly Gln Ala Val Leu Ala Asn Ser Ser Gln Pro
GCT GTC CTG CGG GGC CAG GCC GTG TTG GCC AAC TCT TCC CAG CCT

        90                                  100
Phe Glu Pro Leu Gln Leu His Met Asp Lys Ala Ile Ser Gly Leu
TTC GAG CCC CTG CAG CTG CAC ATG GAT AAA GCC ATC AGT GGC CTT

                            110
Arg Ser Ile Thr Thr Leu Leu Arg Ala Leu Gly Ala Gln Glu Ala
CGC AGC ATC ACC ACT CTG CTT CGG GCG CTG GGA GCC CAG GAA GCC

        120                                 130
Ile Ser Leu Pro Asp Ala Ala Ser Ala Ala Pro Leu Arg Thr Ile
ATC TCC CTC CCA GAT GCG GCC TCG GCT GCT CCA CTC CGA ACC ATC

                            140
Thr Ala Asp Thr Phe Cys Lys Leu Phe Arg Val Tyr Ser Asn Phe
ACT GCT GAC ACT TTC TGC AAA CTC TTC CGA GTC TAC TCC AAT TTC
```

TABLE V (continued)

```
          150                                          160
Leu Arg Gly Lys Leu Lys Leu Tyr Thr Gly Glu Ala Cys Arg Arg
CTC CGG GGA AAG CTG AAG CTG TAC ACG GGG GAG GCC TGC AGG AGA

        165
Gly Asp Arg OP
GGG GAC AGA TGA CCAGGTGCGTCCAGCTGGGCACATCCACCACCTCCCTCACCAACA

CTGCCTGTGCCACACCCTCCCTCACCACTCCCGAACCCCATCGAGGGGCTCTCAGCTAAG

CGCCAGCCTGTCCCATGGACACTCCAGTGCCAGCAATGACATCTCAGGGGCCAGAGGAAC

TGTCCAGAGCACAACTCTGAGATCTAAGGATGTCGCAGGGCCAACTTGAGGGCCCAGAGC

AGGAAGCATTCAGAGAGCAGCTTTAAACTCAGGAGCAGAGACAATGCAGGGAAAACACCT

GAGCTCACTCGGCCACCTGCAAAATTTGATGCAGGACACGCTTTGGAGGCAATTTACCTG

TTTTTGCACCTACCATCAGGGACAGGATGACTGGAGAACTTAGGTGGCAAGCTGTGACTT

CTCAAGGCCTCACGGGCACTCCCTTGGTGGCAAGAGCCCCCTTGACACTGAGAGAATATT

TTGCAATCTGCAGCAGGAAAAATTACGGACAGGTTTTGGAGGTTGGAGGGTACTTGACAG

GTGTGTGGGGAAGCAGGGCGGTAGGGGTGGAGCTGGGATGCGAGTGAGAACCGTGAAGAC

AGGATGGGGGCTGGCCTCTGGTTCTCGTGGGGTCCAAGCTT
                                    HindIII
```


The polypeptide sequence of Table V may readily be subjected to analysis for the presence of highly hydrophilic regions and/or secondary conformational characteristics indicative of potentially highly immunogenic regions by, e.g., the methods of Hopp, et al., P.N.A.S. (U.S.A.), 78, pp. 3824-3828 (1981) and Kyte et al., J.Mol.Biol., 157, pp. 105-132 (1982) and/or Chou, et al., Biochem., 13, pp. 222-245 (1974) and Advances in Enzymology, 47, pp. 45-47 (1978). Computer-assisted analysis according to the Hopp, et al. method is available by means of a program designated PEP Reference Section 6.7 made available by Intelligenetics, Inc., 124 University Avenue, Palo Alto, California.

EXAMPLE 4

A. Human Genomic Library

A Ch4A phage-borne human fetal liver genomic library prepared according to the procedures of Lawn, et al., Cell, 18, pp. 533-543 (1979) was obtained and maintained for use in a plaque hybridization assay.

B. Plaque Hybridization Procedures For Screening Human Genomic Library

Phage particles were lysed and the DNAs were fixed on filters (50,000 plaques per filter) according to the procedures of Woo, Methods In Enzymology, 68, pp. 389-395 (1979) except for the use of GeneScreen Plus filters (New England Nuclear Catalog No. NEF-976) and NZYAM plates (NaCl, 5g; $MgCl_2$-$6H_2O$, 2 g; NZ-Amine A, 10g; yeast extract, 5g; casamino acids, 2 g; maltose; 2g; and agar, 15g per liter).

The air-dried filters were baked at 80·C for 1 hour and then digested with Proteinase K as described in Example 3, Part B. Prehybridization was carried out with a 1M NaCl - 1% SDS buffer for 55·C for 4 hours or more,

after which the buffer was removed. Hybridization and post-hybridization washings were carried out as described in Example 3, Part B. Both the mixture of 128 20-mer probes designated EPV and the mixture of 128 17-mer probes of Table III (designated the EPQ mixture) were employed. Hybridization was carried out at 48°C using the EPV probe mixture. EPQ probe mixture hybridization was carried out at 46°C -- 4 degrees below the lowest calculated Td for members of the mixture. Removal of the hybridized probe for rehybridization was accomplished by boiling with 1 x SSC - 0.1% SDS for two minutes. Autoradiography of the filters revealed three positive clones (reactive with both probe mixtures) among the 1,500,000 phage plaques screened. Verification of the positive clones as being EPO-encoding was obtained through DNA sequencing and electron micrographic visualization of heteroduplex formation with the monkey cDNA of Example 3. This procedure also gave evidence of multiple introns in the genomic DNA sequence.

EXAMPLE 5

Nucleotide sequence analysis of one of the positive clones (designated λhE1) was carried out and results obtained to date are set out in Table VI.

TABLE VI

```
AAGCTTCTGGGCTTCCAGACCCAGCTACTTTGCGGAACTCAGCAACCCAGGCATCTCTGAGTCTCCGCCCA

AGACCGGGATGCCCCCCAGGGGAGGTGTCCGGGAGCCCAGCCTTTCCCAGATAGCACGCTCCGCCAGTCCC

AAGGGTGCGCAACCGGCTGCACTCCCCTCCCGCGACCCAGGGCCCGGGAGCAGCCCCCATGACCCACACGC

ACGTCTGCAGCAGCCCCGCTCACGCCCCGGCGAGCCTCAACCCAGGCGTCCTGCCCCTGCTCTGACCCCGG

GTGGCCCCTACCCCTGGCGACCCCTCACGCACACAGCCTCTCCCCCACCCCCACCCGCGCACGCACACATG

CAGATAACAGCCCCGACCCCCGGCCAGAGCCGXAGAGTCCCTGGGCCACCCCGGCCGCTCGCCTGCCGCTG

CGCCGCACCGCGCTGTCCTCCCGGAGCCGGACCGGGGCCACCGCGCCCXGCTCTGCTCCGACACCGCGCCC

CTTGGACAGCCGCCCTCTCCTCTAGGCCCGTGGGGCTGGCCCTGCACCGCCGAGCTTCCCGGGATGAGGXX

                                                 -27         -24
                                                 Met Gly Val His
CCCGGTGACCGGCGCGCCCCAAGTCGCTGAGGGACCCCGGCCAAGCGCGGAG ATG GGG GTG CAC G

GTGAGTACTCGCGGGCTGGGCGCTCCCGGCGGCCGGGTTCCTGTTTGAGCGGGGATTTAGCGCCCCGGCT
```

TABLE VI (cont'd.)

```
ATTGGCCAAGAGGTGGCTGGGTTCAAGGACCGGCGACTTGTCAAGGACCCCGGAAGGGGGAGGGGGGTGGG

GCAGCCTCCACGTGCCGCGGGGACTTGGGGGAGTTCTTGGGGATGGCAAAAACCTGGCCTGTTGAGGGGCA

CAGTTTGGGGTTGGGGAGGAGGTTTGGGGTTCTGCTGTGCAGTTGTGTCGTTGTCAGTGTCTCG[I.S.]

TTGCACACGCACAGATCAATAAGCCAGAGGCAGCACCTGAGTGCTTGCATGGTTGGGACAGGAAGGACGAG

CTGGGGCAGAGACGTGGGGATGAAGGAAGCTGTCCTTCCACAGCCACCCTTCTCCCCCCCCGCCTGACTCT

                          -23         -20
                          Glu Cys Pro Ala Trp Leu Trp Leu Leu Leu Ser Leu
CAGCCTGGCTATCTGTTCTAG      AA TGT CCT GCC TGG CTG TGG CTT CTC CTG TCC CTG

    -10                                   -1  +1
Leu Ser Leu Pro Leu Gly Leu Pro Val Leu Gly Ala Pro Pro Arg Leu Ile Cys
CTG TCG CTC CCT CTG GGC CTC CCA GTC CTG GGC GCC CCA CCA CGC CTC ATC TGT

        10                                   20                *
Asp Ser Arg Val Leu Glu Arg Tyr Leu Leu Glu Ala Lys Glu Ala Glu Asn Ile
GAC AGC CGA GTC CTG GAG AGG TAC CTC TTG GAG GCC AAG GAG GCC GAG AAT ATC

 26
Thr
ACG GTGAGACCCCTTCCCCAGCACATTCCACAGAACTCACGCTCAGGGCTTCAGGGAACTCCTCCCAGAT

CCAGGAACCTGGCACTTGGTTTGGGGTGGAGTTGGGAAGCTAGACACTGCCCCCCTACATAAGAATAAGTC
```


TABLE VI (cont'd.)

```
TGGTGGCCCCAAACCATACCTGAAACTAGGCAAGGAGCAAAGCCAGCAGATCCTACGCCTGTGGGCCAGGG

                                                      27          30
                                                      Thr Gly Cys Ala Glu
CCAGAGCCTTCAGGGACCCTTGACTCCCCGGGCTGTGTGCATTTCAG       ACG GGC TGT GCT GAA

                            *       40
His Cys Ser Leu Asn Glu Asn Ile Thr Val Pro Asp Thr Lys Val Asn Phe Tyr
CAC TGC AGC TTG AAT GAG AAT ATC ACT GTC CCA GAC ACC AAA GTT AAT TTC TAT

 50                  55
Ala Trp Lys Arg Met Glu
GCC TGG AAG AGG ATG GAG GTGAGTTCCTTTTTTTTTTTTTTTCCTTTCTTTTGGAGAATCTCATT

TGCGAGCCTGATTTTGGATGAAAGGGAGAATGATCGGGGGAAAGGTAAAATGGAGCAGCAGAGATGAGGCT

GCCTGGGCGCAGAGGCTCACGTCTATAATCCCAGGCTGAGATGGCCGAGATGGGAGAATTGCTTGAGCCCT

GGAGTTTCAGACCAACCTAGGCAGCATAGTGAGATCCCCCATCTCTACAAACATTTAAAAAAATTAGTCAG

GTGAAGTGGTGCATGGTGGTAGTCCCAGATATTTGGAAGGCTGAGGCGGGAGGATCGCTTGAGCCCAGGAA

TTTGAGGCTGCAGTGAGCTGTGATCACACCACTGCACTCCAGCCTCAGTGACAGAGTGAGGCCCTGTCTCA
```

TABLE VI (cont'd.)

AAAAAGAAAAGAAAAAAAGAAAAATAATGAGGGCTGTATGGAATACATTCATTATTCATTCACTCACTCACT

CACTCATTCATTCATTCATTCATTCAACAAGTCTTATTGCATACCTTCTGTTTGCTCAGCTTGGTGCTTGG

GGCTGCTGAGGGGCAGGAGGGAGAGGGTGACATGGGTCAGCTCGACTCCCAGAGTCCACTCCCTGTAG

56 60 70
Val Gly Gln Gln Ala Val Glu Val Trp Gln Gly Leu Ala Leu Leu Ser Glu Ala
GTC GGG CAG CAG GCC GTA GAA GTC TGG CAG GGC CTG GCC CTG CTG TCG GAA GCT

80 * 90
Val Leu Arg Gly Gln Ala Leu Leu Val Asn Ser Ser Gln Pro Trp Glu Pro Leu
GTC CTG CGG GGC CAG GCC CTG TTG GTC AAC TCT TCC CAG CCG TGG GAG CCC CTG

100
Gln Leu His Val Asp Lys Ala Val Ser Gly Leu Arg Ser Leu Thr Thr Leu Leu
CAG CTG CAT GTG GAT AAA GCC GTC AGT GGC CTT CGC AGC CTC ACC ACT CTG CTT

110 115
Arg Ala Leu Gly Ala Gln
CGG GCT CTG GGA GCC CAG GTGAGTAGGAGCGGACACTTCTGCTTGCCCTTTCTGTAAGAAGGGGA

GAAGGGTCTTGCTAAGGAGTACAGGAACTGTCCGTATTCCTTCCCTTTCTGTGGCACTGCAGCGACCTCCT

116 120
Lys Glu Ala Ile Ser Pro Pro Asp Ala Ala Ser Ala Ala
GTTTTCTCCTTGGCAG AAG GAA GCC ATC TCC CCT CCA GAT GCG GCC TCA GCT GCT

- 46 -

0148605

TABLE VI (cont'd.)

130 140
Pro Leu Arg Thr Ile Thr Ala Asp Thr Phe Arg Lys Leu Phe Arg Val Tyr Ser
CCA CTC CGA ACA ATC ACT GCT GAC ACT TTC CGC AAA CTC TTC CGA GTC TAC TCC

150 160
Asn Phe Leu Arg Gly Lys Leu Lys Leu Tyr Thr Gly Glu Ala Cys Arg Thr Gly
AAT TTC CTC CGG GGA AAG CTG AAG CTG TAC ACA GGG GAG GCC TGC AGG ACA GGG

166
Asp Arg OP
GAC AGA TGA CCAGGTGTGTCCACCTGGGCATATCCACCACCTCCCTCACCAACATTGCTTGTGCCACA

CCCTCCCCCGCCACTCCTGAACCCCGTCGAGGGGCTCTCAGCTCAGCGCCAGCCTGTCCCATGGACACTCC

AGTGCCAGCAATGACATCTCAGGGGCCAGAGGAACTGTCCAGAGAGCAACTCTGAGATCTAAGGATGTCAC

AGGGCCAACTTGAAGGGCCCAGAGCAGGAAGCATTCAGAGAGCAGCTTTAAACTCAGGGACAGAGCCATGC

TGGGAAGACGCCTGAGCTCACTCGGCACCCTGCAAAATTTGATGCCAGGACACGCTTTGGAGGCGATTTAC

CTGTTTTCGCACCTACCATCAGGGACAGGATGACCTGGAGAACTTAGGTGGCAAGCTGTGACTTCTCCAGG

TCTCACGGGCATGGGCACTCCCTTGGTGGCAAGAGCCCCCTTGACACCGGGGTGGTGGGAACCATGAAGAC

AXGATXGGGGCTGGCCTCTGGCTCTCATGGGGTCCAAGTTTTGTGTATTCTCAACCTATTGACAGACTGAA

ACACAATATGAC

- 47 -

0148605

In Table VI, the initial continuous DNA sequence designates a top strand of 620 bases in what is apparently an untranslated sequence immediately preceding a translated portion of the human EPO gene. More specifically, the sequence appears to comprise the 5' end of the gene which leads up to a translated DNA region coding for the first four amino acids (-27 through -24) of a leader sequence ("presequence"). Four base pairs in the sequence prior to that encoding the beginning of the leader have not yet been unambiguously determined and are therefore designated by an "X". There then follows an intron of about 639 base pairs (439 base pairs of which have been sequenced and the remaining 200 base pairs of which are designated "I.S.") and immediately preceding a codon for glutamine which has been designated as residue -23 of the translated polypeptide. The exon sequence immediately following is seen to code for amino acid residues through an alanine residue (designated as the +1 residue of the amino acid sequence of mature human EPO) to the codon specifying threonine at position +26, whereupon there follows a second intron consisting of 256 bases as specifically designated. Following this intron is an exon sequence for amino acid residues 27 through 55 and thereafter a third intron comprising 612 base pairs commences. The subsequent exon codes for residues 56 through 115 of human EPO and there then commences a fourth intron of 134 bases as specified. Following the fourth intron is an exon coding for residue Nos. 116 through 166 and a "stop" codon (TGA). Finally, Table VI identifies a sequence of 568 base pairs in what appears to be an untranslated 3' region of the human EPO gene, two base pairs of which ("X") have not yet been unambiguously sequenced.

Table VI thus serves to identify the primary structural conformation (amino acid sequence) of mature human EPO as including 166 specified amino acid residues

(estimated M.W. = 18,399). Also revealed in the Table is the DNA sequence coding for a 27 residue leader sequence along with 5' and 3' DNA sequences which may be significant to promoter/operator functions of the human gene operon. Sites for potential glycosylation of the mature human EPO polypeptide are designated in the Table by asterisks. It is worthy of note that the specific amino acid sequence of Table VI likely constitutes that of a naturally occurring allelic form of human erythropoietin. Support for this position is found in the results of continued efforts at sequencing of urinary isolates of human erythropoietin which provided the finding that a significant number of erythropoietin molecules therin have a methionine at residue 126 as opposed to a serine as shown in the Table.

Table VII, below, illustrates the extent of polypeptide sequence homology between human and monkey EPO. In the upper continuous line of the Table, single letter designations are employed to represent the deduced translated polypeptide sequences of human EPO commencing with residue -27 and the lower continuous line shows the deduced polypeptide sequence of monkey EPO commencing at assigned residue number -27. Asterisks are employed to highlight the sequence homologies. It should be noted that the deduced human and monkey EPO sequences reveal an "additional" lysine (K) residue at (human) position 116. Cross-reference to Table VI indicates that this residue is at the margin of a putative mRNA splice junction in the genomic sequence. Presence of the lysine residue in the human polypeptide sequence was further verified by sequencing of a cDNA human sequence clone prepared from mRNA isolated from COS-1 cells transformed with the human genomic DNA in Example 7, <u>infra</u>.

TABLE VII

Comparison of Human and Monkey EPO Polypeptides

```
                -20          -10          +1         10          20          30          40
Human    MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTK
         **************** ******** ************************* * ** * * *************
Monkey   MGVHECPAWLWLLLSLVSLPLGLPVPGAPPRLICDSRVLERYLLEAKEAENVTMGCSESCSLNENITVPDTK

             50         60          70          80          90         100         110
Human    VNFYAWKRMEVGQQAVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKE
         ********************************** * ***** ****** *** ***** ********** *
Monkey   VNFYAWKRMEVGQQAVEVWQGLALLSEAVLRGQAVLANSSQPFEPLQLHMDKAISGLRSITTLLRALGAQ-E

         120          130         140          150          160
Human    AISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRTGDR
         *** ***************** *********************** ***
Monkey   AISLPDAASAAPLRTITADTFCKLFRVYSNFLRGKLKLYTGEACRRGDR
```


EXAMPLE 6

The expression system selected for initial attempts at microbial synthesis of isolatable quantities of EPO polypeptide material coded for by the monkey cDNA provided by the procedures of Example 3 was one involving mammalian host cells (i.e., COS-1 cells, A.T.C.C. No. CRL-1650). The cells were transfected with a "shuttle" vector capable of autonomous replication in E.coli host (by virtue of the presence of pBR322-derived DNA) and the mammalian hosts (by virtue of the presence of SV40 virus-derived DNA).

More specifically, an expression vector was constructed according to the following procedures. The plasmid clone 83 provided in Example 3 was amplified in E.coli and the approximately 1.4kb monkey EPO-encoding DNA was isolated by EcoRI and HindIII digestion. Separately isolated was an approximately 4.0 kb, HindIII/SalI fragment from pBR322. An approximately 30 bp, EcoRI/SalI "linker" fragment was obtained from M13mp10 RF DNA (P and L Laboratories). This linker included, in series, an EcoRI sticky end, followed by SstI, SmaI, BamHI and XbaI recognition sites and a SalI sticky end. The above three fragments were ligated to provide an approximately 5.4 kb intermediate plasmid ("pERS") wherein the EPO DNA was flanked on one side by a "bank" of useful restriction endonuclease recognition sites. pERS was then digested with HindIII and SalI to yield the EPO DNA and the EcoRI to SalI (M13mp10) linker. The 1.4 kb fragment was ligated with an approximately 4.0 kb BamHI/SalI of pBR322 and another M13mp10 HindIII/BamHI RF fragment linker also having approximately 30 bp. The M13 linker fragment was characterized by a HindIII sticky end, followed by PstI, SalI, XbaI recognition sites and a BamHI sticky end. The ligation product was, again, a useful intermediate plasmid ("pBR-EPO") including the EPO DNA flanked on both sides by banks of restriction site.

The vector chosen for expression of the EPO DNA in COS-1 cells ("pDSVL1") had previously been constructed to allow for selection and autonomous replication in E.coli. These characteristics are provided by the origin of replication and Ampicillin resistance gene DNA sequences present in the region spanning nucleotides 2448 through 4362 of pBR322. This sequence was structurally modified by the addition of a linker providing a HindIII recognition immediately adjacent nucleotide 2448 prior to incorporation into the vector. Among the selected vector's other useful properties was the capacity to autonomously replicate in COS-1 cells and the presence of a viral promoter sequence functional in mammalian cells. These characteristics are provided by the origin of replication DNA sequence and "late gene" viral promoter DNA sequence present in the 342 bp sequence spanning nucleotide numbers 5171 through 270 of the SV40 genome. A unique restriction site (BamHI) was provided in the vector and immediately adjacent the viral promoter sequence through use of a commercially available linker sequence (Collaborative Research). Also incorporated in the vector was a 237 base pair sequence (derived as nucleotide numbers 2553 through 2770 of SV40) containing the "late gene" viral mRNA polyadenylation signal (commonly referred to as a transcription terminator). This fragment was positioned in the vector in the proper orientation vis-a-vis the "late gene" viral promoter via the unique BamHI site. Also present in the vector was another mammalian gene at a location not material to potential transcription of a gene inserted at the unique BamHI site, between the viral promoter and terminator sequences. [The mammalian gene comprised an approximately 2,500 bp mouse dihydrofolate reductase (DHFR) minigene isolated from plasmid pMG-1 as in Gasser, et al., P.N.A.S. (U.S.A.), 79, pp. 6522-6526, (1982).] Again, the major operative components of plasmid pDSVL1 comprise nucleotides 2448 through 4362 of pBR322 along with nucleotides 5171 through 270 (342bp) and 2553 through 2770 (237bp) of SV40 DNA.

Following procedures described, e.g., in Maniatis, et al., supra, the EPO-encoding DNA was isolated from plasmid pBR-EPO as a BamHI fragment and ligated into plasmid pDSVL1 cut with BamHI. Restriction enzyme analysis was employed to confirm insertion of the EPO gene in the correct orientation in two of the resulting cloned vectors (duplicate vectors H and L). See Figure 2, illustrating plasmid pDSVL-MkE. Vectors with EPO genes in the wrong orientation were saved for use as negative controls in transfection experiments designed to determine EPO expression levels in hosts transformed with vectors having EPO DNA in the correct orientation.

Vectors H, L, F, X and G were combined with carrier DNA (mouse liver and spleen DNA) were employed to transfect duplicate 60mm plates by calcium phosphate microprecipitate methods. Duplicate 60 mm plates were also transfected with carrier DNA as a "mock" transformation negative control. After five days all culture media were tested for the presence of polypeptides possessing the immunological properties of naturally-occurring EPO.

EXAMPLE 7

A. Initial EPO Expression System Involving COS-1 Cells

The system selected for initial attempts at microbial synthesis of isolatable quantities of human EPO polypeptide material coded for by the human genomic DNA EPO clone, also involved expression in mammalian host cells (i.e., COS-1 cells, A.T.C.C. No. CRL-1650). The

human EPO gene was first sub-cloned into a "shuttle" vector which is capable of autonomous replication in both E.coli hosts (by virtue of the presence of pBR322 derived DNA) and in the mammalian cell line COS-1 (by virtue of the presence of SV40 virus derived DNA). The shuttle vector, containing the EPO gene, was then transfected into COS-1 cells. EPO polypeptide material was produced in the transfected cells and secreted into the cell culture media.

More specifically, an expression vector was constructed according to the following procedures. DNA isolated from lambda clone λhE1, containing the human genomic EPO gene, was digested with BamHI and HindIII restriction endonucleases, and a 5.6 Kb DNA fragment known to contain the entire EPO gene was isolated. This fragment was mixed and ligated with the bacterial plasmid pUC8 (Bethesda Research Laboratories, Inc.) which had been similarly digested, creating the intermediate plasmid "pUC8-HuE", providing a convenient source of this restriction fragment.

The vector chosen for expression of the EPO DNA in COS-1 cells (pSV4SEt) had previously been constructed. Plasmid pSV4SEt contained DNA sequences allowing selection and autonomous replication in E.coli. These characteristics are provided by the origin of replication and Ampicillin resistance gene DNA sequences present in the region spanning nucleotides 2448 through 4362 of the bacterial plasmid pBR322. This sequence was structurally modified by the addition of a linker providing a HindIII recognition site immediately adjacent to nucleotide 2448. Plasmid pSV4SEt was also capable of autonomous replication in COS-1 cells. This characteristic was provided by a 342 bp fragment containing the SV40 virus origin of replication (nucleotide numbers 5171 through 270). This fragment had been modified by the addition of a linker providing an EcoRl recognition site adjacent to

nucleotide 270 and a linker providing a SalI recognition site adjacent nucleotide 5171. A 1061 bp fragment of SV40 was also present in this vector (nucleotide numbers 1711 through 2772 plus a linker providing a SalI recognition site next to nucleotide number 2772). Within this fragment was an unique BamHI recognition sequence. In summary, plasmid pSV4SEt contained unique BamHI and HindIII recognition sites, allowing insertion of the human EPO gene, sequences allowing replication and selection in E.coli, and sequences allowing replication in COS-1 cells.

In order to insert the EPO gene into pSV4SEt, plasmid pUC8-HuE was digested with BamHl and HindIII restriction endonucleases and the 5.6 kb EPO encoding DNA fragment isolated. pSV4SEt was also digested with BamHl and HindIII and the major 2513 bp fragment isolated (preserving all necessary functions). These fragments were mixed and ligated, creating the final vector "pSVgHuEPO". (See, Figure 3.) This vector was propagated in E.coli and vector DNA isolated. Restriction enzyme analysis was employed to confirm insertion of the EPO gene.

Plasmid pSVgHuEPO DNA was used to express human EPO polypeptide material in COS-1 cells. More specifically, pSVgHuEPO DNA was combined with carrier DNA and transfected into triplicate 60 mm plates of COS-1 cells. As a control, carrier DNA alone was also transfected into COS-1 cells. Cell culture media were sampled five and seven days later and tested for the presence of polypeptides possessing the immunological properties of naturally occurring human EPO.

B. Second EPO Expression System Involving COS-1 Cells

Still another system was designed to provide improved production of human EPO polypeptide material

coded by the human genomic DNA EPO clone in COS-1 cells (A.T.C.C. No. CRL-1650).

In the immediately preceding system, EPO was expressed in COS-1 cells using its own promoter which is within the 5.6 Kb BamHI to HindIII restriction fragment. In the following construction, the EPO gene is altered so that it is expressed using the SV40 late promoter.

More specifically, the cloned 5.6 Kb BamHI to HindIII genomic human EPO restriction fragment was modified by the following procedures. Plasmid pUC8-HuE, as described above, was cleaved with BamHI and with BstEII restriction endonucleases. BstEII cleaves within the 5.6 Kb EPO gene at a position which is 44 base pairs 5' to the initiating ATG coding for the pre-peptide and approximately 680 base pairs 3' to the HindIII restriction site. The approximately 4900 base pair fragment was isolated. A synthetic linker DNA fragment, containing SalI and BstEII sticky ends and an internal BamHI recognition site was synthesized and purified. The two fragments were mixed and ligated with plasmid pBR322 which had been cut with SalI and BamHI to produce the intermediate plasmid pBRgHE. The genomic human EPO gene can be isolated therefrom as a 4900 base pair BamHI digestion fragment carrying the complete structural gene with a single ATG 44 base pairs 3' to BamHI site adjacent the amino terminal coding region.

This fragment was isolated and inserted as a BamHI fragment into BamHI cleaved expression vector plasmid pDSVL1 (described in Example 6). The resulting plasmid, pSVLgHuEPO, as illustrated in Figure 4, was used to express EPO polypeptide material from COS-1 cells, as described in Examples 6 and 7A.

EXAMPLE 8

Culture media from growth of the six transfected COS-1 cultures of Example 6 were analyzed by radioimmunoassay according to the procedures set forth in Example 2, Part B. Each sample was assayed at 250, 125, 50, and 25 microliter aliquot levels. Supernatants from growth of cells mock transfected or transfected with vectors having incorrect EPO gene orientation were unambiguously negative for EPO immunoreactivity. For each sample of the two supernatants derived from growth of COS-1 cells transfected with vectors (H and L) having the EPO DNA in the correct orientation, the % inhibition of ^{125}I-EPO binding to antibody ranged from 72 to 88%, which places all values at the top of the standard curve. The exact concentration of EPO in the culture supernatant could not then reliably be estimated. A quite conservative estimate of 300 mU/ml was made, however, from the value calculation of the largest aliquot size (250 microliter).

A representative culture fluid according to Example 6 and five and seven day culture fluids obtained according to Example 7A were tested in the RIA in order to compare activity of recombinant monkey and human EPO materials to a naturally-occurring human EPO standard and the results are set out in graphic form in Figure 1. Briefly, the results expectedly revealed that the recombinant monkey EPO significantly competed for anti-human EPO antibody although it was not able to completely inhibit binding under the test conditions. The maximum percent inhibition values for recombinant human EPO, however, closely approximated those of the human EPO standard. The parallel nature of the dose response curves suggests immunological identity of the sequences (epitopes) in common. Prior estimates of monkey EPO in culture fluids were re-evaluated at these higher dilution levels and were found to range from 2.91 to 3.12 U/ml. Estimated human EPO production levels were correspondingly set at 392 mU/ml for the five-day growth sample

and 567 mU/ml for the seven day growth sample. Estimated monkey EPO production levels in the Example 7B expression system were on the same order or better.

EXAMPLE 9

Culture fluids prepared according to Examples 6 and 7 were subjected to an *in vitro* assay for EPO activity according to the procedure of Goldwasser, et al., Endocrinology, 97, 2, pp. 315-323 (1975). Estimated monkey EPO values for culture fluids tested ranged from 3.2 to 4.3 U/ml. Human EPO culture fluids were also active in this *in vitro* assay and, further, this activity could be neutralized by anti-EPO antibody. The recombinant monkey EPO culture fluids according to Example 6 were also subjected to an assay for *in vivo* biological activity according to the general procedures of Cotes, et al., Nature, 191, pp. 1065-1067 (1961) and Hammond, et al., Ann.N.Y.Acad.Sci., 149, pp. 516-527 (1968) and activity levels ranged from 0.94 to 1.24 U/ml.

EXAMPLE 10

In the previous examples, recombinant monkey or human EPO material was produced from vectors used to transfect COS-1 cells. These vectors replicate in COS-1 cells due to the presence of SV40 T antigen within the cell and an SV40 origin of replication on the vectors. Though these vectors produce useful quantities of EPO in COS-1 cells, expression is only transient (7 to 14 days) due to the eventual loss of the vector. Additionally, only a small percentage of COS-1 became productively transfected with the vectors. The present example describes expression systems employing Chinese hamster ovary (CHO) $DHFR^-$ cells and the selectable marker, DHFR. [For discussion of related expression systems, see

U.S. Letters Patent No. 4,399,216 and European Patent Applications 117058, 117059 and 117060, all published August 29, 1984.]

CHO $DHFR^-$ cells (DuX-B11) CHO K1 cells, Urlaub, et al., Proc. Nat. Acad. Sci. (U.S.A.), Vol. 77, 4461 (1980) lack the enzyme dihydrofolate reductase (DHFR) due to mutations in the structural genes and therefore require the presence of glycine, hypoxanthine, and thymidine in the culture media. Plasmids pDSVL-MkE (Example 6) or pDSVL-gHuEPO (Example 7B) were transfected along with carrier DNA into CHO $DHFR^-$ cells growing in media containing hypoxanthine, thymidine, and glycine in 60 mm culture plates. Plasmid pSVgHuEPO (Example 7A) was mixed with the plasmid pMG2 containing a mouse dihydrofolate reductase gene cloned into the bacterial plasmid vector pBR322 (per Gasser, et al., *supra*.) The plasmid mixture and carrier DNA was transfected into CHO $DHFR^-$ cells. (Cells which acquire one plasmid will generally also acquire a second plasmid). After three days, the cells were dispersed by trypsinization into several 100 mm culture plates in media lacking hypoxanthine and thymidine. Only those cells which have been stably transformed with the DHFR gene, and thereby the EPO gene, survive in this media. After 7-21 days, colonies of surviving cells became apparent. These transformant colonies, after dispersion by trypsinization can be continuously propagated in media lacking hypoxanthine and thymidine, creating new cell strains (e.g., CHO pDSVL-MkEPO, CHO pSVgHuEPO, CHO-pDSVL-gHuEPO).

Culture fluids from the above cell strains were tested in the RIA for the presence of recombinant monkey or human EPO. Media for strain CHO pDSVL-MkEPO contained EPO with immunological properties like that obtained from COS-1 cells transfected with plasmid pDSVL-MkEPO. A representative 65 hour culture fluid contained monkey EPO at 0.60 U/ml.

Culture fluids from CHO pSVgHuEPO and CHO pDSVL-gHuEPO contained recombinant human EPO with immunological properties like that obtained with COS-1 cells transfected with plasmid pSVgHuEPO or pDSVL-gHuEPO. A representative 3 day culture fluid from CHO pSVgHuEPO contained 2.99 U/ml of human EPO and a 5.5 day sample from CHO pDSVL-gHuEPO had 18.2 U/ml of human EPO as measured by the RIA.

The quantity of EPO produced by the cell strains described above can be increased by gene amplification giving new cell strains of greater productivity. The enzyme dihydrofolate reductase (DHFR) which is the product coded for by the DHFR gene can be inhibited by the drug methotrexate (MTX). More specifically, cells propagated in media lacking hypoxanthine and thymidine are inhibited or killed by MTX. Under the appropriate conditions, (e.g., minimal concentrations of MTX) cells resistant to and able to grow in MTX can be obtained. These cells are found to be resistent to MTX due to an amplification of the number of their DHFR genes, resulting in increased production of DHFR enzyme. The surviving cells can, in turn, be treated with increasing concentrations of MTX, resulting in cell strains containing greater numbers of DHFR genes. "Passenger genes" (e.g., EPO) carried on the expression vector along with the DHFR gene or transformed with the DHFR gene are frequently found also to be increased in their gene copy number.

As examples of practice of this amplification system, cell strain CHO pDSVL-MkE was subjected to increasing MTX concentrations (0 nM, 30 nM and 100 nM). Representative 65-hour culture media samples from each amplification step were assayed by RIA and determined to contain 0.60, 2.45 and 6.10 U/ml, respectively. Cell strain CHO pDSVL-gHuEPO was subjected to a series of increasing MTX concentrations of 30 nM, 50 nM, 100 nM,

200 nM, 1 μM, and 5 μM MTX. A representative 3-day culture media sample from the 100 nM MTX step contained human EPO at 3089 ± 129 u/ml as judged by RIA. Representative 48 hour cultural medium samples from the 100 nM and 1 μM MTX steps contained, respectively, human EPO at 466 and 1352 U/ml as judged by RIA (average of triplicate assays). In these procedures, 1×10^6 cells were plated in 5 ml of media in 60 mm culture dishes. Twenty-four hours later the media were removed and replaced with 5 ml of serum-free media (high glucose DMEM supplemented with 0.1 mM non-essential amino acids and L-glutamine). EPO was allowed to accumulate for 48 hours in the serum-free media. The media was collected for RIA assay and the cells were trypsinized and counted. The average RIA values of 467 U/ml and 1352 U/ml for cells grown at 100 nM and 1 μM MTX, respectively, provided actual yields of 2335 U/plate and 6750 U/plate. The average cell numbers per plate were 1.94×10^6 and 3.12×10^6 cells, respectively. The effective production rates for these culture conditions were thus 1264 and 2167 U/10^6 cells/48 hours.

The cells in the cultures described immediately above are a genetically heterogeneous population. Standard screening procedures are being employed in an attempt to isolate genetically hemogeneous clones with the highest production capacity. See, Section A, Part 2, of "Points to Consider in the Characterization of Cell Lines Used to Produce Biologics", June 1, 1984, Office of Biologics Research Review, Center for Drugs and Biologics, U.S. Food and Drug Administration.

The productivity of the EPO producing CHO cell lines described above can be improved by appropriate cell culture techniques. The propagation of mammalian cells in culture generally requires the presence of serum in the growth media. A method for production of erythropoietin from CHO cells in media that does not contain

serum greatly facilitates the purification of erythropoietin from the culture medium. The method described below is capable of economically producing erythropoietin in serum-free media in large quantities sufficient for production.

Strain CHO pDSVL-gHuEPO cells, grown in standard cell culture conditions, are used to seed spinner cell culture flasks. The cells are propagated as a suspension cell line in the spinner cell culture flask in media consisting of a 50-50 mixture of high glucose DMEM and Ham's F12 supplemented with 5% fetal calf serum, L-glutamine, Penicillin and Streptomycin, 0.05 mM non-essential amino acids and the appropriate concentration of methotrexate. Suspension cell culture allows the EPO-producing CHO cells to be expanded easily to large volumes. CHO cells, grown in suspension, are used to seed roller bottles at an initial seeding density of 1.5×10^7 viable cells per 850 cm^2 roller bottle in 200 ml of media. The cells are allowed to grow to confluency as an adherent cell line over a three-day period. The media used for this phase of the growth is the same as used for growth in suspension. At the end of the three-day growth period, the serum containing media is removed and replaced with 100 ml of serum-free media; 50-50 mixture of high glucose DMEM and Ham's F12 supplemented with 0.05 mM non-essential amino acids and L-glutamine. The roller bottles are returned to the roller bottle incubator for a period of 1-3 hours and the media again is removed and replaced with 100 ml of fresh serum-free media. The 1-3 hour incubation of the serum-free media reduces the concentration of contaminating serum proteins. The roller bottles are returned to the incubator for seven days during which erythropoietin accumulates in the serum-free culture media. At the end of the seven-day production phase, the conditioned media is removed and replaced with fresh serum-free medium for a second

production cycle. As an example of the practice of this production system, a representative seven-day, serum-free media sample contained human erythropoietin at 3892±409 U/ml as judged by the RIA. Based on an estimated cell density of 0.9 to 1.8×10^5 cells/cm^2, each 850 cm^2 roller bottle contained from 0.75 to 1.5×10^8 cells and thus the rate of production of EPO in the 7-day, 100 ml culture was 750 to 1470 U/10^6 cells/48 hours.

Culture fluids from cell strain CHO pDSVL-MkEPO carried in 10 nM MTX were subjected to RIA <u>in vitro</u> and <u>in vivo</u> EPO activity assays. The conditioned media sample contained 41.2 ± 1.4 U/ml of MkEPO as measured by the RIA, 41.2 ± 0.064 U/ml as measured by the <u>in vitro</u> biological activity assay and 42.5 ± 5 U/ml as measured by the <u>in vivo</u> biological activity assay. Amino acid sequencing of polypeptide products revealed the presence of EPO products, a principle species having 3 residues of the "leader" sequence adjacent the putative amino terminal alanine. Whether this is the result of incorrect membranc processing of the polypeptide in CHO cells or reflects a difference in structure of the amino terminus of monkey EPO <u>vis-a-vis</u> human EPO, is presently unknown.

Culture fluids from cell strain CHO pDSVL-gHuEPO were subjected to the three assays. A 5.5 day sample contained recombinant human EPO in the media at a level of 18.2 U/ml by RIA assay, 15.8 ± 4.6 U/ml by <u>in vitro</u> assay and 16.8 ± 3.0 U/ml by <u>in vivo</u> assay.

Culture fluid from CHO pDSVL-gHuEPO cells prepared amplified by stepwise 100 nM MTX were subjected to the three assays. A 3.0 day sample contained recombinant human EPO at a level of 3089 ± 129 U/ml by RIA, 2589 ± 71.5 U/ml by <u>in vitro</u> assay, and 2040 ± 160 U/ml by <u>in vivo</u> assay. Amino acid sequencing of this product reveals an amino terminal corresponding to that designated in Table VI.

Cell conditioned media from CHO cells transfected with plasmid pDSVL-MkE in 10 nM MTX were pooled,

and the MTX dialyzed out over several days, resulting in media with an EPO activity of 221 ± 5.1 U/ml (EPO-CCM). To determine the _in vivo_ effect of the EPO-CCM upon hematocrit levels in normal Balb/C mice, the following experiment was conducted. Cell conditioned media from untransfected CHO cells (CCM) and EPO-CCM were adjusted with PBS. CCM was used for the control group (3 mice) and two dose levels of EPO-CCM -- 4 units per injection and 44 units per injection -- were employed for the experimental groups (2 mice/group). Over the course of 5 weeks, the seven mice were injected intraperitoneally, 3 times per week. After the eighth injection, average hematocrit values for the control group were determined to be 50.4%; for the 4U group, 55.1%; and, for the 44U group, 67.9%.

Mammalian cell expression products may be readily recovered in substantially purified form from culture media using HPLC (C_4) employing an ethanol gradient, preferably at pH7.

A preliminary attempt was made to characterize recombinant glycoprotein products from conditioned medium of COS-1 and CHO cell expression of the human EPO gene in comparison to human urinary EPO isolates using both Western blot analysis and SDS-PAGE. These studies indicated that the CHO-produced EPO material had a somewhat higher molecular weight than the COS-1 expression product which, in turn, was slightly larger than the pooled source human urinary extract. All products were somewhat heterogeneous. Neuraminidase enzyme treatment to remove sialic acid resulted in COS-1 and CHO recombinent products of approximately equal molecular weight which were both nonetheless larger than the resulting asialo human urinary extract. Endoglycosidase F enzyme (EC 3.2.1) treatment of the recombinant CHO product and the urinary extract product (to totally remove carbohydrate from

both) resulted in substantially homogeneous products having essentially identical molecular weight characteristics.

Purified human urinary EPO and a recombinant, CHO cell-produced, EPO according to the invention were subjected to carbohydrate analysis according to the procedure of Ledeen, et al. _Methods in Enzymology_, _83(Part D)_, 139-191 (1982) as modified through use of the hydrolysis procedures of Nesser, et al., _Anal.Biochem._, _142_, 58-67 (1984). Experimentally determined carbohydrate constitution values (expressed as molar ratios of carbohydrate in the product) for the urinary isolate were as follows: Hexoses, 1.73; N-acetylglucosamine, 1; N-acetylneuraminic acid, 0.93; Fucose, 0; and N-acetylgalactosamine, 0. Corresponding values for the recombinant product (derived from CHO pDSVL-gHuEPO 3-day culture media at 100 nM MTX) were as follows: Hexoses, 15.09; N-acetylglucosamine, 1; N-acetylneuraminic acid, 0.998; Fucose, 0; and N-acetylgalactosamine, 0. These findings are consistent with the Western blot and SDS-PAGE analysis described above.

Glycoprotein products provided by the present invention are thus comprehensive of products having a primary structural conformation sufficiently duplicative of that of a naturally-occurring erythropoietin to allow possession of one or more of the biological properties thereof and having an average carbohydrate composition which differs from that of naturally-occurring erythropoietin.

EXAMPLE 11

The present example relates to the total manufacture by assembly of nucleotide bases of two structural genes encoding the human species EPO sequence of Table VI and incorporating, respectively "preferred" codons for expression in _E.coli_ and yeast (_S.cerevisiae_) cells.

Also described is the construction of genes encoding analogs of human EPO. Briefly stated, the protocol employed was generally as set out in the previously noted disclosure of Alton, et al. (WO 83/04053). The genes were designed for initial assembly of component oligonucleotides into multiple duplexes which, in turn, were assembled into three discrete sections. These sections were designed for ready amplification and, upon removal from the amplification system, could be assembled sequentially or through a multiple fragment ligation in a suitable expression vector.

Tables VIII through XIV below illustrate the design and assembly of a manufactured gene encoding a human EPO translation product lacking any leader or presequence but including an initial methionine residue at position -1. Moreoever, the gene incorporated in substantial part *E.coli* preference codons and the construction was therefore referred to as the "ECEPO" gene.

TABLE VIII

ECEPO SECTION 1 OLIGONUCLEOTIDES

1. AATTCTAGAAACCATGAGGGTAATAAAATA
2. CCATTATTTTATTACCCTCATGGTTTCTAG
3. ATGGCTCCGCCGCGTCTGATCTGCGAC
4. CTCGAGTCGCAGATCAGACGCGGCGGAG
5. TCGAGAGTTCTGGAACGTTACCTGCTG
6. CTTCCAGCAGGTAACGTTCCAGAACT
7. GAAGCTAAAGAAGCTGAAAACATC
8. GTGGTGATGTTTTCAGCTTCTTTAG
9. ACCACTGGTTGTGCTGAACACTGTTC
10. CAAAGAACAGTGTTCAGCACAACCA
11. TTTGAACGAAAACATTACGGTACCG
12. GATCCGGTACCGTAATGTTTTCGTT

TABLE IX

ECEPO SECTION 1

```
    XbaI
EcoRI              1                              3
AATTCTAG AAACCATGAG GGTAATAAAA TAATGGCTCC GCCGCGTCTG
    GATC TTTGGTACTC CCATTATTTT ATTACCGAGG CGGCGCAGAC
                        2                         4

                5
ATCTGCGACT CGAGAGTTCT GGAACGTTAC CTGCTGGAAG CTAAAGAAGC
TAGACGCTGA GCTCTCAAGA CCTTGCAATG GACGACCTTC GATTTCTTCG
                          6

        7             9                         11
TGAAAACATC ACCACTGGTT GTGCTGAACA CTGTTCTTTG AACGAAAACA
ACTTTTGTAG TGGTGACCAA CACGACTTGT GACAAGAAAC TTGCTTTTGT
         8                 10

    KpnI    BamHI
TTACGGTACC G
AATGCCATGG CCTAG
  12
```

TABLE X
ECEPO SECTION 2 OLIGONUCLEOTIDES

1.	AATTCGGTACCAGACACCAAGGT
2.	GTTAACCTTGGTGTCTGGTACCG
3.	TAACTTCTACGCTTGGAAACGTAT
4.	TTCCATACGTTTCCAAGCGTAGAA
5.	GGAAGTTGGTCAACAAGCAGTTGAAGT
6.	CCAAACTTCAACTGCTTGTTGACCAAC
7.	TTGGCAGGGTCTGGCACTGCTGAGCG
8.	GCCTCGCTCAGCAGTGCCAGACCCTG
9.	AGGCTGTACTGCGTGGCCAGGCA
10.	GCAGTGCCTGGCCACGCAGTACA
11.	CTGCTGGTAAACTCCTCTCAGCCGT
12.	TTCCCACGGCTGAGAGGAGTTTACCA
13.	GGGAACCGCTGCAGCTGCATGTTGAC
14.	GCTTTGTCAACATGCAGCTGCAGCGG
15.	AAAGCAGTATCTGGCCTGAGATCTG
16.	GATCCAGATCTCAGGCCAGATACT

TABLE XI
ECEPO SECTION 2

```
EcoRI  KpnI           1                         3
A ATTCGGTACC AGACACCAAG GTTAACTTCT ACGCTTGGAA ACGTATGGAA
      GCCATGG TCTGTGGTTC CAATTGAAGA TGCGAACCTT TGCATACCTT
                 2                         4

       5                           7
GTTGGTCAAC AAGCAGTTGA AGTTTGGCAG GGTCTGGCAC TGCTGAGCGA
CAACCAGTTG TTCGTCAACT TCAAACCGTC CCAGACCGTG ACGACTCGCT
     6                                   8

                9                          11
GGCTGTACTG CGTGGCCAGG CACTGCTGGT AAACTCCTCT CAGCCGTGGG
CCGACATGAC GCACCGGTCC GTGACGACCA TTTGAGGAGA GTCGGCACCC
                10                    12

     13                              15        BglII    BamHI
AACCGCTGCA GCTGCATGTT GACAAAGCAG TATCTGGCCT GAGATCTG
TTGGCGACGT CGACGTACAA CTGTTTCGTC ATAGACCGGA CTCTAGACCTAC
     14                                  16
```

TABLE XII

ECEPO SECTION 3

1.	GATCCAGATCTCTGACTACTCTGC
2.	ACGCAGCAGAGTAGTCAGAGATCTG
3.	TGCGTGCTCTGGGTGCACAGAAAGAGG
4.	GATAGCCTCTTTCTGTGCACCCAGAGC
5.	CTATCTCTCCGCCGGATGCTGCATCT
6.	CAGCAGATGCAGCATCCGGCGGAGA
7.	GCTGCACCGCTGCGTACCATCACTG
8.	ATCAGCAGTGATGGTACGCAGCGGTG
9.	CTGATACCTTCCGCAAACTGTTTCG
10.	ATACACGAAACAGTTTGCGGAAGGT
11.	TGTATACTCTAACTTCCTGCGTGGTA
12.	CAGTTTACCACGCAGGAAGTTAGAGT
13.	AACTGAAACTGTATACTGGCGAAGC
14.	GGCATGCTTCGCCAGTATACAGTTT
15.	ATGCCGTACTGGTGACCGCTAATAG
16.	TCGACTATTAGCGGTCACCAGTAC

TABLE XIII

ECEPO SECTION 3

```
BamHI BglII
GA TCCAGATCTCTG
     GTCTAGAGAC

     1          3                     5
ACTACTCTGC TGCGTGCTCT GGGTGCACAG AAAGAGGCTA TCTCTCCGCC
TGATGAGACG ACGCACGAGA CCCACGTGTC TTTCTCCGAT AGAGAGGCGG
       2                     4

                       7                     9
GGATGCTGCA TCTGCTGCAC CGCTGCGTAC CATCACTGCT GATACCTTCC
CCTACGACGT AGACGACGTG GCGACGCATG GTAGTGACGA CTATGGAAGG
     6                   8

                      11                    13
GCAAACTGTT TCGTGTATAC TCTAACTTCC TGCGTGGTAA ACTGAAACTG
CGTTTGACAA AGCACATATG AGATTGAAGG ACGCACCATT TGACTTTGAC
    10                        12

                      15                SalI
TATACTGGCG AAGCATGCCG TACTGGTGAC CGCTAATAG
ATATGACCGC TTCGTACGGC ATGACCACTG GCGATTATC  AGCT
    14                      16
```

TABLE XIV

ECEPO GENE

```
                                        -1  1
XbaI                                    MetAla
CTAG AAACCATGAG GGTAATAAAA TAATGGCTCC GCCGCGTCTG
         TTTGGTACTC CCATTATTTT ATTACCGAGG CGGCGCAGAC

ATCTGCGACT CGAGAGTTCT GGAACGTTAC CTGCTGGAAG CTAAAGAAGC
TAGACGCTGA GCTCTCAAGA CCTTGCAATG GACGACCTTC GATTTCTTCG

TGAAAACATC ACCACTGGTT GTGCTGAACA CTGTTCTTTG AACGAAAACA
ACTTTTGTAG TGGTGACCAA CACGACTTGT GACAAGAAAC TTGCTTTTGT

TTACGGTACC AGACACCAAG GTTAACTTCT ACGCTTGGAA ACGTATGGAA
AATGCCATGG TCTGTGGTTC CAATTGAAGA TGCGAACCTT TGCATACCTT

GTTGGTCAAC AAGCAGTTGA AGTTTGGCAG GGTCTGGCAC TGCTGAGCGA
CAACCAGTTG TTCGTCAACT TCAAACCGTC CCAGACCGTG ACGACTCGCT

GGCTGTACTG CGTGGCCAGG CACTGCTGGT AAACTCCTCT CAGCCGTGGG
CCGACATGAC GCACCGGTCC GTGACGACCA TTTGAGGAGA GTCGGCACCC

AACCGCTGCA GCTGCATGTT GACAAAGCAG TATCTGGCCT GAGATCTCTG
TTGGCGACGT CGACGTACAA CTGTTTCGTC ATAGACCGGA CTCTAGAGAC

ACTACTCTGC TGCGTGCTCT GGGTGCACAG AAAGAGGCTA TCTCTCCGCC
TGATGAGACG ACGCACGAGA CCCACGTGTC TTTCTCCGAT AGAGAGGCGG

GGATGCTGCA TCTGCTGCAC CGCTGCGTAC CATCACTGCT GATACCTTCC
CCTACGACGT AGACGACGTG GCGACGCATG GTAGTGACGA CTATGGAAGG

GCAAACTGTT TCGTGTATAC TCTAACTTCC TGCGTGGTAA ACTGAAACTG
CGTTTGACAA AGCACATATG AGATTGAAGG ACGCACCATT TGACTTTGAC

                                            SalI
TATACTGGCG AAGCATGCCG TACTGGTGAC CGCTAATAG
ATATGACCGC TTCGTACGGC ATGACCACTG GCGATTATCA GCT
```


More particularly, Table VIII illustrates oligonucleotides employed to generate the Section 1 of the ECEPO gene encoding amino terminal residues of the human species polypeptide. Oligonucleotides were assembled into duplexes (1 and 2, 3 and 4, etc.) and the duplexes were then ligated to provide ECEPO Section 1 as in Table IX. Note that the assembled section includes respective terminal EcoRI and BamHI sticky ends, that "downstream" of the EcoRI sticky end is a XbaI restriction enzyme recognition site; and that "upstream" of the BamHI sticky end is a KpnI recognition site. Section 1 could readily be amplified using the M13 phage vector employed for verification of sequence of the section. Some difficulties were encountered in isolating the section as an XbaI/KpnI fragment from RF DNA generated in E.coli, likely due to methylation of the KpnI recognition site bases within the host. Single-stranded phage DNA was therefore isolated and rendered into double-stranded form in vitro by primer extension and the desired double-stranded fragment was thereafter readily isolated.

ECEPO gene Sections 2 and 3 (Tables XI and XIII) were constructed in a similar manner from the oligonucleotides of Tables X and XII, respectively. Each section was amplified in the M13 vector employed for sequence verification and was isolated from phage DNA. As is apparent from Table XI, ECEPO Section 2 was constructed with EcoRI and BamHI sticky ends and could be isolated as a KpnI/BglII fragment. Similarly, ECEPO Section 3 was prepared with BamHI and SalI sticky ends and could be isolated from phage RF DNA as a BglII/SalI fragment. The three sections thus prepared can readily be assembled into a continuous DNA sequence (Table XIV) encoding the entire human species EPO polypeptide with an amino terminal methionine codon (ATG) for E.coli translation initiation. Note also that "upstream" of the initial ATG is a series of base pairs substantially

duplicating the ribosome binding site sequence of the highly expressed OMP-f gene of E.coli.

Any suitable expression vector may be employed to carry the ECEPO. The particular vector chosen for expression of the ECEPO gene as the "temperature sensitive" plasmid pCFM536 -- a derivative of plasmid pCFM414 (A.T.C.C. 40076) -- as described in co-pending U.S. Patent Application Serial No. 636,727, filed August 6, 1984, by Charles F. Morris. More specifically, pCFM536 was digested with XbaI and HindIII; the large fragment was isolated and employed in a two-part ligation with the ECEPO gene. Sections 1 (XbaI/KpnI), 2 (KpnI/BglII) and 3 (BglII/SalI) had previously been assembled in the correct order in M13 and the EPO gene was isolated therefrom as a single XbaI/HindIII fragment. This fragment included a portion of the polylinker from M13 mp9 phage spanning the SalI to HindIII sites therein. Control of expression in the resulting expression plasmid, p536, was by means of a lambda P_L promoter, which itself may be under control of the C_{I857} repressor gene (such as provided in E.coli strain K12ΔHtrp).

The manufactured ECEPO gene above may be variously modified to encode erythropoietin analogs such as [Asn^2, des-Pro^2 through Ile^6]hEPO and [His^7]hEPO, as described below.

A. [Asn^2, des-Pro^2 through Ile^6]hEPO

Plasmid 536 carrying the ECEPO manufactured gene of Table XIV as a XbaI to HindIII insert was digested with HindIII and XhoI. The latter endonuclease cuts the ECEPO gene at a unique, 6 base pair recognition site spanning the last base of the codon encoding Asp^8 through the second base of the Arg^{10} codon. A XbaI/XhoI "linker" sequence was manufactured having the following sequence:


```
XbaI       +1  2   7   8   9
        Met Ala Asn Cys Asp   XhoI
5'-CTAG ATG GCT AAT TGC GAC-3'
    3'-TAC CGA TTA ACG CTG AGCT-5'
```

The XbaI/XhoI linker and the XhoI/HindIII ECEPO gene sequence fragment were inserted into the large fragment resulting from XbaI and HindIII digestion of plasmid pCFM526 -- a derivative of plasmid pCFM414 (A.T.C.C. 40076) -- as described in co-pending U.S. Patent Application Serial No. 636,727, filed August 6, 1984, by Charles F. Morris, to generate a plasmid-borne DNA sequence encoding E.coli expression of the Met^{-1} form of the desired analog.

B. [His^7]hEPO

Plasmid 536 was digested with HindIII and XhoI as in part A above. A XbaI/XhoI linker was manufactured having the following sequence:

```
XbaI     +1  2   3   4   5   6   7   8   9     XhoI
     Met Ala Pro Pro Arg Leu Ile His Asp
5'-CTAG ATG GCT CCG CCA CGT CTG ATC CAT GAC-3'
    3'-TAC CGA GGC GGT GCA GAC TAG GTA CTG  AGCT-5'
```

The linker and the XhoI/HindIII ECEPO sequence fragment were then inserted into pCFM526 to generate a plasmid-borne DNA sequence encoding E.coli expression of the Met^{-1} form of the desired analog.

Construction of a manufactured gene ("SCEPO") incorporating yeast preference codons is as described in the following Tables XV through XXI. As was the case with the ECEPO gene, the entire construction involved formation of three sets of oligonucleotides (Tables XV, XVII and XIX) which were formed into duplexes and assembled into sections (Tables XVI, XVIII and XX). Note that synthesis was facilitated in part by use of some sub-optimal codons in both the SCEPO and ECEPO construc-

tions, i.e., oligonucleotides 7-12 of Section 1 of both genes were identical, as were oligonucleotides 1-6 of Section 2 in each gene.

TABLE XV

SCEPO SECTION 1 OLIGONUCLEOTIDES

1. AATTCAAGCTTGGATAAAAGAGCT
2. GTGGAGCTCTTTTATCCAAGCTTG
3. CCACCAAGATTGATCTGTGACTC
4. TCTCGAGTCACAGATCAATCTTG
5. GAGAGTTTTGGAAAGATACTTGTTG
6. CTTCCAACAAGTATCTTTCCAAAAC
7. GAAGCTAAAGAAGCTGAAAACATC
8. GTGGTGATGTTTTCAGCTTCTTTAG
9. ACCACTGGTTGTGCTGAACACTGTTC
10. CAAAGAACAGTGTTCAGCACAACCA
11. TTTGAACGAAAACATTACGGTACCG
12. GATCCGGTACCGTAATGTTTTCGTT

TABLE XVI

SCEPO SECTION 1

```
EcoRI  HindIII 1
AATTCA AGCTTGGATA
    GT TCGAACCTAT
               2

                     3
AAAGAGCTCC ACCAAGATTG ATCTGTGACT CGAGAGTTTT
TTTCTCGAGG TGGTTCTAAC TAGACACTGA GCTCTCAAAA
                           4

    5                           7
GGAAAGATAC TTGTTGGAAG CTAAAGAAGC TGAAAACATC ACCACTGGTT
CCTTTCTATG AACAACCTTC GATTTCTTCG ACTTTTGTAG TGGTGACCAA
    6                            8

       9                       11        KpnI     BamHI
GTGCTGAACA CTGTTCTTTG AACGAAAACA TTACGGTACC G
CACGACTTGT GACAAGAAAC TTGCTTTTGT AATGCCATGG CCTAG
                                   12
```

TABLE XVII

SCEPO SECTION 2 OLIGONUCLEOTIDES

1. AATTCGGTACCAGACACCAAGGT
2. GTTAACCTTGGTGTCTGGTACCG
3. TAACTTCTACGCTTGGAAACGTAT
4. TTCCATACGTTTCCAAGCGTAGAA
5. GGAAGTTGGTCAACAAGCAGTTGAAGT
6. CCAAACTTCAACTGCTTGTTGACCAAC
7. TTGGCAAGGTTTGGCCTTGTTATCTG
8. GCTTCAGATAACAAGGCCAAACCTTG
9. AAGCTGTTTTGAGAGGTCAAGCCT
10. AACAAGGCTTGACCTCTCAAAACA
11. TGTTGGTTAACTCTTCTCAACCATGGG
12. TGGTTCCCATGGTTGAGAAGAGTTAACC
13. AACCATTGCAATTGCACGTCGAT
14. CTTTATCGACGTGCAATTGCAA
15. AAAGCCGTCTCTGGTTTGAGATCTG
16. GATCCAGATCTCAAACCAGAGACGG

TABLE XVIII

SCEPO SECTION 2

```
        KpnI
EcoRI    1
A ATTCGGTACC AGACACCAAG
     GCCATGG TCTGTGGTTC
             2

         3                                5
GTTAACTTCT ACGCTTGGAA ACGTATGGAA GTTGGTCAAC AAGCTGTTGA
CAATTGAAGA TGCGAACCTT TGCATACCTT CAACCAGTTG TTCGACAACT
              4                                6

             7                         9
AGTTTGGCAA GGTTTGGCCT TGTTATCTGA AGCTGTTTTG AGAGGTCAAG
TCAAACCGTT CCAAACCGGA ACAATAGACT TCGACAAAAC TCTCCAGTTC
                 8                              10

               11                          13
CCTTGTTGGT TAACTCTTCT CAACCATGGG AACCATTGCA ATTGCACGTC
GGAACAACCA ATTGAGAAGA GTTGGTACCC TTGGTAACGT TAACGTGCAG
                  12                             14

            15        BglII   BamHI
GATAAAGCCG TCTCTGGTTT GAGATCTG
CTATTTCGGC AGAGACCAAA CTCTAGACCTA G
              16
```

TABLE XIX

SCEPO SECTION 3 OLIGONUCLEOTIDES

1. GATCCAGATCTTTGACTACTTTGTT
2. TCTCAACAAAGTAGTCAAAGATCTG
3. GAGAGCTTTGGGTGCTCAAAAGGAAG
4. ATGGCTTCCTTTTGAGCACCCAAAGC
5. CCATTTCCCCACCAGACGCTGCTT
6. GCAGAAGCAGCGTCTGGTGGGGAA
7. CTGCCGCTCCATTGAGAACCATC
8. CAGTGATGGTTCTCAATGGAGCG
9. ACTGCTGATACCTTCAGAAAGTT
10. GAATAACTTTCTGAAGGTATCAG
11. ATTCAGAGTTTACTCCAACTTCT
12. CTCAAGAAGTTGGAGTAAACTCT
13. TGAGAGGTAAATTGAAGTTGTACAC
14. ACCGGTGTACAACTTCAATTTACCT
15. CGGTGAAGCCTGTAGAACTGGT
16. CTGTCACCAGTTCTACAGGCTTC
17. GACAGATAAGCCCGACTGATAA
18. GTTGTTATCAGTCGGGCTTAT
19. CAACAGTGTAGATGTAACAAAG
20. TCGACTTTGTTACATCTACACT

TABLE XX

SCEPO SECTION 3

```
BamHI  BglII  1
GATC CAGATCTTTG ACTACTTTGT TGAGAGCTTT
     GTCTAGAAAC TGATGAAACA ACTCTCGAAA
                    2

  3                       5
GGGTGCTCAA AAGGAAGCCA TTTCCCCACC AGACGCTGCT TCTGCCGCTC
CCCACGAGTT TTCCTTCGGT AAAGGGGTGG TCTGCGACGA AGACGGCGAG
  4                         6

   7                 9                       11
CATTGAGAAC CATCACTGCT GATACCTTCA GAAAGTTATT CAGAGTTTAC
GTAACTCTTG GTAGTGACGA CTATGGAAGT CTTTCAATAA GTCTCAAATG
    8                     10                        12

                      13                       15
TCCAACTTCT TGAGAGGTAA ATTGAAGTTG TACACCGGTG AAGCCTGTAG
AGGTTGAAGA ACTCTCCATT TAACTTCAAC ATGTGGCCAC TTCGGACATC
                        14                        16

              17                       19
AACTGGTGAC AGATAAGCCC GACTGATAAC AACAGTGTAG
TTGACCACTG TCTATTCGGG CTGACTATTG TTGTCACATC
                  18

             SalI
ATGTAACAAA G
TACATTGTTT CAGCT
   20
```

TABLE XXI

SCEPO GENE

```
              -1 +1
HindIII       ArgAla
AGCTTGGATA AAAGAGCTCC ACCAAGATTG ATCTGTGACT CGAGAGTTTT
    ACCTAT TTTCTCGAGG TGGTTCTAAC TAGACACTGA GCTCTCAAAA

GGAAAGATAC TTGTTGGAAG CTAAAGAAGC TGAAAACATC ACCACTGGTT
CCTTTCTATG AACAACCTTC GATTTCTTCG ACTTTTGTAG TGGTGACCAA

GTGCTGAACA CTGTTCTTTG AACGAAAACA TTACGGTACC AGACACCAAG
CACGACTTGT GACAAGAAAC TTGCTTTTGT AATGCCATGG TCTGTGGTTC

GTTAACTTCT ACGCTTGGAA ACGTATGGAA GTTGGTCAAC AAGCTGTTGA
CAATTGAAGA TGCGAACCTT TGCATACCTT CAACCAGTTG TTCGACAACT

AGTTTGGCAA GGTTTGGCCT TGTTATCTGA AGCTGTTTTG AGAGGTCAAG
TCAAACCGTT CCAAACCGGA ACAATAGACT TCGACAAAAC TCTCCAGTTC

CCTTGTTGGT TAACTCTTCT CAACCATGGG AACCATTGCA ATTGCACGTC
GGAACAACCA ATTGAGAAGA GTTGGTACCC TTGGTAACGT TAACGTGCAG

GATAAAGCCG TCTCTGGTTT GAGATCTTTG ACTACTTTGT TGAGAGCTTT
CTATTTCGGC AGAGACCAAA CTCTAGAAAC TGATGAAACA ACTCTCGAAA

GGGTGCTCAA AAGGAAGCCA TTTCCCCACC AGACGCTGCT TCTGCCGCTC
CCCACGAGTT TTCCTTCGGT AAAGGGGTGG TCTGCGACGA AGACGGCGAG

CATTGAGAAC CATCACTGCT GATACCTTCA GAAAGTTATT CAGAGTTTAC
GTAACTCTTG GTAGTGACGA CTATGGAAGT CTTTCAATAA GTCTCAAATG

TCCAACTTCT TGAGAGGTAA ATTGAAGTTG TACACCGGTG AAGCCTGTAG
AGGTTGAAGA ACTCTCCATT TAACTTCAAC ATGTGGCCAC TTCGGACATC

AACTGGTGAC AGATAAGCCC GACTGATAAC AACAGTGTAG
TTGACCACTG TCTATTCGGG CTGACTATTG TTGTCACATC

            SalI
ATGTAACAAA G
TACATTGTTT CAGCT
```


The assembled SCEPO sections were sequenced in M13 and Sections 1, 2 and 3 were isolatable from the phage as HindIII/KpnI, KpnI/BglII, and BglII/SalI fragments.

The presently preferred expression system for SCEPO gene products is a secretion system based on *S.cerevisiae* α-factor secretion, as described in co-pending U.S. Patent Application Serial No. 487,753, filed April 22, 1983, by Grant A. Bitter, published October 31, 1984 as European Patent Application 0 123,294. Briefly put, the system involves constructions wherein DNA encoding the leader sequence of the yeast α-factor gene product is positioned immediately 5' to the coding region of the exogenous gene to be expressed. As a result, the gene product translated includes a leader or signal sequence which is "processed off" by an endogenous yeast enzyme in the course of secretion of the remainder of the product. Because the construction makes use of the α-factor translation initiation (ATG) codon, there was no need to provide such a codon at the -1 position of the SCEPO gene. As may be noted from Table XXI, the alanine (+1) encoding sequence is preceded by a linker sequence allowing for direct insertion into a plasmid including the DNA for the first 80 residues of the α-factor leader following the α-factor promoter. The specific preferred construction for SCEPO gene expression involved a four-part ligation including the above-noted SCEPO section fragments and the large fragment of HindIII/SalI digestion of plasmid pαC3. From the resulting plasmid pαC3/SCEPO, the α-factor promoter and leader sequence and SCEPO gene were isolated by digestion with BamHI and ligated into BamHI digested plasmid pYE to form expression plasmid pYE/SCEPO.

EXAMPLE 12

The present example relates to expression of

recombinant products of the manufactured ECEPO and SCEPO genes within the expression systems of Example 11.

In use of the expression system designed for use of E.coli host cells, plasmid p536 of Example 11 was transformed into AM7 E.coli cells previously transformed with a suitable plasmid, pMW1, harboring a C_{I857} gene. Cultures of cells in LB broth (Ampicillin 50 µg/ml and kanamycin 5 µg/ml, preferably with 10 mM $MgSO_4$) were maintained at 28°C and upon growth of cells in culture to $O.D._{600} = 0.1$, EPO expression was induced by raising the culture temperature to 42°C. Cells grown to about 40 O.D. provided EPO production (as estimated by gel) of about 5 mg/OD liter.

Cells were harvested, lysed, broken with French Press (10,000 psi) and treated with lysozyme and NP-40 detergent. The pellet resulting from 24,000 xg centrifugation was solubilized with guanidine HCl and subjected to further purification in a single step by means of C_4 (Vydac) Reverse Phase HPLC (EtOH, 0-80%, 50 mM NH_4Ac, pH 4.5). Protein sequencing revealed the product to be greater than 95% pure and the products obtained revealed two different amino terminals, A-P-P-R... and P-P-R... in a relative quantitative ratio of about 3 to 1. This latter observation of hEPO and [des Ala^1]hEPO products indicates that amino terminal "processing" within the host cells serves to remove the terminal methionine and in some instances the initial alanine. Radioimmunoassay activity for the isolates was at a level of 150,000 to 160,000 U/mg; in vitro assay activity was at a level of 30,000 to 62,000 U/mg; and in vivo assay activity ranged from about 120 to 720 U/mg. (Cf., human urinary isolate standard of 70,000 U/mg in each assay.) The dose response curve for the recombinant product in the in vivo assay differed markedly from that of the human urinary EPO standard.

The EPO analog plasmids formed in parts A and B of Example 11 were each transformed into pMW1-transformed AM7 E.coli cells and the cells were cultured as above. Purified isolates were tested in both RIA and in vitro assays. RIA and in vitro assay values for [Asn^2, des-Pro^2 through Ile^6]hEPO expression products were approximately 11,000 U/mg and 6,000 U/mg protein, respectively, while the assay values for [His^7]hEPO were about 41,000 U/mg and 14,000 U/mg protein, respectively, indicating that the analog products were from one-fourth to one-tenth as "active" as the "parent" expression product in the assays.

In the expression system designed for use of S.cerevisiae host cells, plasmid pYE/SCEPO was transformed into two different strains, YSDP4 (genotype α pep4-3 trpl) and RK81 (genotype αα pep4-3 trpl). Transformed YSDP4 hosts were grown in SD medium (Methods in Yeast Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., p. 62 (1983) supplemented with casamino acids at 0.5%, pH 6.5 at 30°C. Media harvested when the cells had been grown to 36 O.D. contained EPO products at levels of about 244 U/ml (97 µg/OD liter by RIA). Transformed RK81 cells grown to either 6.5 O.D. or 60 O.D. provided media with EPO concentrations of about 80-90 U/ml (34 µg/OD liter by RIA). Preliminary analyses reveal significant heterogeneity in products produced by the expression system, likely to be due to variations in glycosylation of proteins expressed, and relatively high mannose content of the associated carbohydrate.

Plasmids PαC3 and pYE in HB101 E.coli cells were deposited in accordance with the Rules of Practice of the U.S. Patent Office on September 27, 1984, with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland, under deposit numbers A.T.C.C. 39881 and A.T.C.C. 39882, respectively. Plasmids pCFM526 in AM7 cells, pCFM536 in JM103 cells, and pMW1 in JM103

cells were likewise deposited on November 21, 1984 as A.T.C.C. 39932, 39934, and 39933, respectively. Saccharomyces cerevisiae strains YSPD4 and RK81 were deposited on November 21, 1984 as A.T.C.C. 20734 and 20733, respectively.

It should be readily apparent from consideration of the above illustrative examples that numerous exceptionally valuable products and processes are provided by the present invention in its many aspects.

Polypeptides provided by the invention are conspicuously useful materials, whether they are microbially expressed products or synthetic products, the primary, secondary or tertiary structural conformation of which was first made known by the present invention.

As previously indicated, recombinant-produced and synthetic products of the invention share, to varying degrees, the in vitro biological activity of EPO isolates from natural sources and consequently are projected to have utility as substitutes for EPO isolates in culture media employed for growth of erythropoietic cells in culture. Similarly, to the extent that polypeptide products of the invention share the in vivo activity of natural EPO isolates they are conspicuously suitable for use in erythropoietin therapy procedures practiced on mammals, including humans, to develop any or all of the effects herefore attributed in vivo to EPO, e.g., stimulation of reticulocyte response, development of ferrokinetic effects (such as plasma iron turnover effects and marrow transit time effects), erythrocyte mass changes, stimulation of hemoglobin C synthesis (see, Eschbach, et al., supra) and, as indicated in Example 10, increasing hematocrit levels in mammals. Included within the class of humans treatable with products of the invention are patients generally requiring blood transfusions and including trauma victims, surgical patients, renal disease patients including dialysis patients, and

patients with a variety of blood composition affecting disorders, such as hemophilia, sickle cell disease, physiologic anemias, and the like. The minimization of the need for transfusion therapy through use of EPO therapy can be expected to result in reduced transmission of infectious agents. Products of the invention, by virtue of their production by recombinant methods, are expected to be free of pyrogens, natural inhibitory substances, and the like, and are thus likely to provide enhanced overall effectiveness in therapeutic processes vis-a-vis naturally derived products. Erythropoietin therapy with products of the present invention is also expected to be useful in the enhancement of oxygen carrying capacity of individuals encountering hypoxic environmental conditions and possibly in providing beneficial cardiovascular effects.

A preferred method for administration of polypeptide products of the invention is by parenteral (e.g., IV, IM, SC, or IP) routes and the compositions administered would ordinarily include therapeutically effective amounts of product in combination with acceptable diluents, carriers and/or adjuvants. Preliminary pharmacokinetic studies indicate a longer half-life in vivo for monkey EPO products when administered IM rather than IV. Effective dosages are expected to vary substantially depending upon the condition treated but therapeutic doses are presently expected to be in the range of 0.1 (~7U) to 100 (~7000U) µg/kg body weight of the active material. Standard diluents such as human serum albumin are contemplated for pharmaceutical compositions of the invention, as are standard carriers such as saline.

Adjuvant materials suitable for use in compositions of the invention include compounds independently noted for erythropoietic stimulatory effects, such as testosterones, progenitor cell stimulators, insulin-like growth factor, prostaglandins, serotonin,

cyclic AMP, prolactin and triiodothyronine, as well as agents generally employed in treatment of aplastic anemia, such as methenolene, stanozolol and nandrolone [see, e.g., Resegotti, et al., Panminerva Medica, 23, 243-248 (1981); McGonigle, et al., Kidney Int., 25(2), 437-444 (1984); Pavlovic-Kantera, et al., Expt.Hematol., 8(Supp. 8), 283-291 (1980); and Kurtz, FEBS Letters, 14a(1), 105-108 (1982)]. Also contemplated as adjuvants are substances reported to enhance the effects of, or synergize, erythropoietin or asialo-EPO, such as the adrenergic agonists, thyroid hormones, androgens and BPA [see, Dunn, "Current Concepts in Erythropoiesis", John Wiley and Sons (Chichester, England, 1983); Weiland, et al., Blut, 44(3), 173-175 (1982); Kalmanti, Kidney Int., 22, 383-391 (1982); Shahidi, New.Eng.J.Med., 289, 72-80 (1973); Fisher, et al., Steroids, 30(6), 833-845 (1977); Urabe, et al., J.Exp.Med., 149, 1314-1325 (1979); and Billat, et al., Expt.Hematol., 10(1), 133-140 (1982)] as well as the classes of compounds designated "hepatic erythropoietic factors" [see, Naughton, et al., Acta.Haemat., 69, 171-179 (1983)] and "erythrotropins" [as described by Congote, et al. in Abstract 364, Proceedings 7th International Congress of Endocrinology (Quebec City, Quebec, July 1-7, 1984); Congote, Biochem.Biophys.Res.Comm., 115(2), 447-483 (1983) and Congote, Anal.Biochem., 140, 428-433 (1984)] and "erythrogenins" [as described in Rothman, et al., J.Surg.Oncol., 20, 105-108 (1982)]. Preliminary screenings designed to measure erythropoietic responses of ex-hypoxic polycythemic mice pre-treated with either 5-α-dihydrotestosterone or nandrolone and then given erythropoietin of the present invention have generated equivocal results.

Diagnostic uses of polypeptides of the invention are similarly extensive and include use in labelled and unlablled forms in a variety of immunoassay techniques

including RIA's, ELISA's and the like, as well as a variety of in vitro and in vivo activity assays. See, e.g., Dunn, et al., Expt.Hematol., 11(7), 590-600 (1983); Gibson, et al., Pathology, 16, 155-156 (1984); Krystal, Expt.Hematol., 11(7), 649-660 (1983); Saito, et al., Jap.J.Med., 23(1), 16-21 (1984); Nathan, et al., New Eng.J.Med., 308(9), 520-522 (1983); and various references pertaining to assays referred to therein. Polypeptides of the invention, including synthetic peptides comprising sequences of residues of EPO first revealed herein, also provide highly useful pure materials for generating polyclonal antibodies and "banks" of monoclonal antibodies specific for differing continuous and discontinuous epitopes of EPO. As one example, preliminary analysis of the amino acid sequences of Table VI in the context of hydropathicity according to Hopp, et al., P.N.A.S. (U.S.A.), 78, pp. 3824-3828 (1981) and of secondary structures according to Chou, et al., Ann.Rev.Biochem., 47, p. 251 (1978) revealed that synthetic peptides duplicative of continuous sequences of residues spanning positions 41-57 inclusive, 116-118 inclusive and 144-166 inclusive are likely to produce a highly antigenic response and generate useful monoclonal and polyclonal antibodies immunoreactive with both the synthetic peptide and the entire protein. Such antibodies are expected to be useful in the detection and affinity purification of EPO and EPO-related products.

Illustratively, the following three synthetic peptides were prepared:

(1) hEPO 41-57, V-P-D-T-K-V-N-F-Y-A-W-K-R-M-E-V-G;

(2) hEPO 116-128, K-E-A-I-S-P-P-D-A-A-S-A-A;

(3) hEPO 144-166, V-Y-S-N-F-L-R-G-K-L-K-L-Y-T-G-E-A-C-R-T-G-D-R.

Preliminary immunization studies employing the above-noted polypeptides have revealed a relatively weak positive response to hEPO 41-57, no appreciable response to hEPO 116-128, and a strong positive resopnse to hEPO 144-166, as measured by capacity of rabbit serum antibodies to immunoprecipitate ^{125}I-labelled human urinary EPO isolates. Preliminary _in vivo_ activity studies on the three peptides revealed no significant activity either alone or in combination.

While the deduced sequences of amino acid residues of mammalian EPO provided by the illustrative examples essentially define the primary structural conformation of mature EPO, it will be understood that the specific sequence of 165 amino acid residues of monkey species EPO in Table V and the 166 residues of human species EPO in Table VI do not limit the scope of useful polypeptides provided by the invention. Comprehended by the present invention are those various naturally-occurring allelic forms of EPO which past research into biologically active mammalian polypeptides such as human γ interferon indicates are likely to exist. (Compare, e.g., the human immune interferon species reported to have an arginine residue at position No. 140 in EPO published application 0 077 670 and the species reported to have glutamine at position No. 140 in Gray, et al., _Nature_, _295_, pp. 503-508 (1982). Both species are characterized as constituting "mature" human γ interferon sequences.) Allelic forms of mature EPO polypeptides may vary from each other and from the sequences of Tables V and VI in terms of length of sequence and/or in terms of deletions, substitutions, insertions or additions of amino acids in the sequence, with consequent potential variations in the capacity for glycosylation. As noted previously, one putative allelic form of human species EPO is believed to include a methionine residue at position 126. Expectedly, naturally-occurring allelic forms

of EPO-encoding DNA genomic and cDNA sequences are also likely to occur which code for the above-noted types of allelic polypeptides or simply employ differing codons for designation of the same polypeptides as specified.

In addition to naturally-occurring allelic forms of mature EPO, the present invention also embraces other "EPO products" such as polypeptide analogs of EPO and fragments of "mature" EPO. Following the procedures of the above-noted published application by Alton, et al. (WO/83/04053) one may readily design and manufacture genes coding for microbial expression of polypeptides having primary conformations which differ from that herein specified for mature EPO in terms of the identity or location of one or more residues (e.g., substitutions, terminal and intermediate additions and deletions). Alternately, modifications of cDNA and genomic EPO genes may be readily accomplished by well-known site-directed mutagenesis techniques and employed to generate analogs and derivatives of EPO. Such EPO products would share at least one of the biological properties of EPO but may differ in others. As examples, projected EPO products of the invention include those which are foreshortened by e.g., deletions [Asn^2, des-Pro^2 through Ile^6]hEPO, [des-Thr^{163} through Arg^{166}]hEPO and "Δ27-55hEPO", the latter having the residues coded for by an entire exon deleted; or which are more stable to hydrolysis (and, therefore, may have more pronounced or longer lasting effects than naturally-occurring EPO); or which have been altered to delete one or more a potential sites for glycosylation (which may result in higher activities for yeast-produced products); or which have one or more cystein residues deleted or replaced by, e.g., histidine or serine residues (such as the analog [His^7]hEPO) and are potentially more easily isolated in active form from microbial systems; or which have one or more tyrosine residues replaced by phenylalanine (such as the analogs

[Phe^{15}]hEPO, [Phe^{49}]hEPO, and [Phe^{145}]hEPO) and may bind more or less readily to EPO receptors on target cells. Also comprehended are polypeptide fragments duplicating only a part of the continuous amino acid sequence or secondary conformations within mature EPO, which fragments may possess one activity of EPO (e.g., receptor binding) and not others (e.g., erythropoietic activity). Especially significant in this regard are those potential fragments of EPO which are elucidated upon consideration of the human genomic DNA sequence of Table VI, i.e., "fragments" of the total continuous EPO sequence which are delineated by intron sequences and which may constitute distinct "domains" of biological activity. It is noteworthy that the absence of in vivo activity for any one or more of the "EPO products" of the invention is not wholly preclusive of therapeutic utility (see, Weiland, et al., supra) or of utility in other contexts, such as in EPO assays or EPO antagonism. Antagonists of erythropoietin may be quite useful in treatment of polycythemias or cases of overproduction of EPO [see, e.g., Adamson, Hosp.Practice, 18(12), 49-57 (1983), and Hellmann, et al., Clin.Lab.Haemat., 5, 335-342 (1983)].

According to another aspect of the present invention, the cloned DNA sequences described herein which encode human and monkey EPO polypeptides are conspicuously valuable for the information which they provide concerning the amino acid sequence of mammalian erythropoietin which has heretofore been unavailable despite decades of analytical processing of isolates of naturally-occurring products. The DNA sequences are also conspicuously valuable as products useful in effecting the large scale microbial synthesis of erthropoietin by a variety of recombinant techniques. Put another way, DNA sequences provided by the invention are useful in generating new and useful viral and circular plasmid DNA vectors, new and useful transformed and transfected

microbial procaryotic and eucaryotic host cells (including bacterial and yeast cells and mammalian cells grown in culture), and new and useful methods for cultured growth of such microbial host cells capable of expression of EPO and EPO products. DNA sequences of the invention are also conspicuously suitable materials for use as labelled probes in isolating EPO and related protein encoding cDNA and genomic DNA sequences of mammalian species other than human and monkey species herein specifically illustrated. The extent to which DNA sequences of the invention will have use in various alternative methods of protein synthesis (e.g., in insect cells) or in genetic therapy in humans and other mammals cannot yet be calculated. DNA sequences of the invention are expected to be useful in developing transgenic mammalian species which may serve as eucaryotic "hosts" for production of erythropoietin and erythropoietin products in quantity. See, generally, Palmiter, et al., Science, 222(4625), 809-814 (1983).

Viewed in this light, therefore, the specific disclosures of the illustrative examples are clearly not intended to be limiting upon the scope of the present invention and numerous modifications and variations are expected to occur to those skilled in the art. As one example, while DNA sequences provided by the illustrative examples include cDNA and genomic DNA sequences, because this application provides amino acid sequence information essential to manufacture of DNA sequence, the invention also comprehends such manufactured DNA sequences as may be constructed based on knowledge of EPO amino acid sequences. These may code for EPO (as in Example 12) as well as for EPO fragments and EPO polypeptide analogs (i.e., "EPO Products") which may share one or more biological properties of naturally-occurring EPO but not share others (or possess others to different degrees).

DNA sequences provided by the present invention are thus seen to comprehend all DNA sequences suitable

for use in securing expression in a procaryotic or eucaryotic host cell of a polypeptide product having at least a part of the primary structural conformation and one or more of the biological properties of erythropoietin, and selected from among: (a) the DNA sequences set out in Tables V and VI; (b) DNA sequences which hybridize to the DNA sequences defined in (a) or fragments thereof; and (c) DNA sequences which, but for the degeneracy of the genetic code, would hybridize to the DNA sequences defined in (a) and (b). It is noteworthly in this regard, for example, that existing allelic monkey and human EPO gene sequences and other mammalian species gene sequences are expected to hybridize to the sequences of Tables V and VI or to fragments thereof. Further, but for the degeneracy of the genetic code, the SCEPO and ECEPO genes and the manufactured or mutagenized cDNA or genomic DNA sequences encoding various EPO fragments and analogs would also hybridize to the above-mentioned DNA sequences. Such hybridizations could readily be carried out under the hybridization conditions described herein with respect to the initial isolation of the monkey and human EPO-encoding DNA or more stringent conditions, if desired to reduce background hybridization.

In a like manner, while the above examples illustrate the invention of microbial expression of EPO products in the context of mammalian cell expression of DNA inserted in a hybrid vector of bacterial plasmid and viral genomic origins, a wide variety of expression systems are within the contemplation of the invention. Conspicuously comprehended are expression systems involving vectors of homogeneous origins applied to a variety of bacterial, yeast and mammlain cells in culture as well as to expression systems not involving vectors (such as calcium phosphate transfection of cells). In

this regard, it will be understood that expression of, e.g., monkey origin DNA in monkey host cells in culture and human host cells in culture, actually constitute instances of "exogenous" DNA expression inasmuch as the EPO DNA whose high level expression is sought would not have its origins in the genome of the host. Expression systems of the invention further contemplate these practices resulting in cytoplasmic formation of EPO products and accumulation of glycosylated and non-glycosylated EPO products in host cell cytoplasm or membrances (e.g., accumulation in bacterial periplasmic spaces) or in culture medium supernatants as above illustrated, or in rather uncommon systems such as P.aeruginosa expression systems (described in Gray, et al., Biotechnology, 2, pp. 161-165 (1984)).

Improved hybridization methodologies of the invention, while illustratively applied above to DNA/DNA hybridization screenings are equally applicable to RNA/RNA and RNA/DNA screening. Mixed probe techniques as herein illustrated generally constitute a number of improvements in hybridization processes allowing for more rapid and reliable polynucleotide isolations. These many individual processing improvements include: improved colony transfer and maintenance procedures; use of nylon-based filters such as GeneScreen and GeneScreen Plus to allow reprobing with same filters and repeated use of the filter, application of novel protease treatments [compared, e.g., to Taub, et al. Anal.Biochem., 126, pp. 222-230 (1982)]; use of very low individual concentrations (on the order of 0.025 picomole) of a large number of mixed probes (e.g., numbers in excess of 32); and, performing hybridization and post-hybridization steps under stringent temperatures closely approaching (i.e., within 4·C and preferably within 2·C away from) the lowest calculated dissocation temperature of any of the mixed probes employed. These improvements combine to

provide results which could not be expected to attend their use. This is amply illustrated by the fact that mixed probe procedures involving 4 times the number of probes ever before reported to have been successfully used in even cDNA screens on messenger RNA species of relatively low abundancy were successfully applied to the isolation of a unique sequence gene in a genomic library screening of 1,500,000 phage plaques. This feat was accomplished essentially concurrently with the publication of the considered opinion of Anderson, et al., supra, that mixed probe screening methods were "...impractical for isolation of mammalian protein genes when corresponding RNA's are unavailable.

The features disclosed in the foregoing description, in the following claims and/or in the accompanying drawings may, both separately and in any combination thereof, be material for realising the invention in diverse forms thereof.

WHAT IS CLAIMED IS:

1. A purified and isolated polypeptide having part or all of the primary structural conformation and one or more of the biological properties of naturally-occurring erythropoietin and characterized by being the product of procaryotic or eucaryotic expression of an exogenous DNA sequence.

2. A polypeptide according to claim 1 further characterized by being free of association with any mammalian protein.

3. A polypeptide according to claim 1 wherein the exogenous DNA sequence is a cDNA sequence.

4. A polypeptide according to claim 1 wherein the exogenous DNA sequence is a manufactured DNA sequence.

5. A polypeptide according to claim 1 wherein the exogenous DNA sequence is a genomic DNA sequence.

6. A polypeptide according to claim 1 wherein the exogenous DNA sequence is carried on an autonomously replicating circular DNA plasmid or viral vector.

7. A polypeptide according to claim 1 possessing part or all of the primary structural conformation of human erythropoietin as set forth in Table VI or any naturally occurring allelic variant thereof.

8. A polypeptide according to claim 1 possessing part or all of the primary structural conformation of monkey erythropoietin as set forth in Table V or any naturally occurring allelic variant thereof.

9. A polypeptide according to claim 1 which has the immunological properties of naturally-occurring erythropoietin.

10. A polypeptide according to claim 1 which has the <u>in vivo</u> biological activity of naturally-occurring erythropoietin.

11. A polypeptide according to claim 1 which has the <u>in vitro</u> biological activity of naturally-occurring erythropoietin.

12. A polypeptide according to claim 1 further characterized by being covalently associated with a detectable label substance.

13. A polypeptide according to claim 12 wherein said detectable label is a radiolabel.

14. A DNA sequence for use in securing expression in a procaryotic or eucaryotic host cell of a polypeptide product having at least a part of the primary structural conformation and one or more of the biological properties of naturally-occurring erythropoietin, said DNA sequence selected from among:

(a) the DNA sequences set out in Tables V and VI or their complementary strands;

(b) DNA sequences which hybridize to the DNA sequences defined in (a) or fragments thereof; and

(c) DNA sequences which, but for the degeneracy of the genetic code, would hybridize to the DNA sequences defined in (a) and (b).

15. A procaryotic or eucaryotic host cell transformed or transfected with a DNA sequence according to claim 14 in a manner allowing the host cell to express said polypeptide product.

16. A polypeptide product of the expression of a DNA sequence of claim 14 in a procaryotic or eucaryotic host.

17. A purified and isolated DNA sequence coding for procaryotic or eucaryotic host expression of a polypeptide having part or all of the primary structural conformation and one or more of the biological properties of erythropoietin.

18. A cDNA sequence according to claim 17.

19. A monkey species erythropoietin coding DNA sequence according to claim 18.

20. A DNA sequence according to claim 19 and including the protein coding region set forth in Table V.

21. A genomic DNA sequence according to claim 17.

22. A human species erythropoietin coding DNA sequence according to claim 21.

23. A DNA sequence according to claim 22 and including the protein coding region set forth in Table VI.

24. A manufactured DNA sequence according to claim 14.

25. A manufactured DNA sequence according to claim 24 and including one or more codons preferred for expression in <u>E.coli</u> cells.

26. A manufactured DNA sequence according to claim 25, coding for expression of human species erythropoietin.

27. A manufactured DNA sequence according to claim 26 including the protein coding region set forth in Table XIV.

28. A manufactured DNA sequence according to claim 24 and including one or more codons preferred for expression in yeast cells.

29. A manufactured DNA sequence according to claim 28, coding for expression of human species erythropoietin.

30. A manufactured DNA sequence according to claim 29 including the protein coding region set forth in Table XXI.

31. A DNA sequence according to claim 17 covalently associated with a detectable label substance.

32. A DNA sequence according to claim 31 wherein the detectable label is a radiolabel.

33. A single-strand DNA sequence according to claim 31.

34. A DNA sequence coding for a polypeptide fragment or polypeptide analog of naturally-occurring erythropoietin.

35. A DNA sequence coding for [Phe^{15}]hEPO, [Phe^{49}]hEPO, [Phe^{145}]hEPO, [His^{7}]hEPO, [Asn^{2} des-Pro^{2} through Ile^{6}]hEPO, [des-Thr^{163} through Arg^{166}]hEPO, or [Δ27-55]hEPO.

36. A DNA sequence according to claim 34 which is a manufactured sequence.

37. A biologically functional circular plasmid or viral DNA vector including a DNA sequence according to either of claims 14, 17, 34 or 35.

38. A procaryotic or eucaryotic host cell stably transformed or transfected with a DNA vector according to claim 37.

39. A polypeptide product of the expression in a procaryotic or eucaryotic host cell of a DNA sequence according to claims 17 or 34.

40. A glycoprotein product having a primary structural conformation sufficiently duplicative of that of a naturally-occurring erythropoietin to allow possession of one or more of the biological properties thereof and having an average carbohydrate composition which differs from that of naturally-occurring erythropoietin.

41. A glycoprotein product having a primary structural conformation sufficiently duplicative of that of a naturally-occurring human erythropoietin to allow possession of one or more of the biological properties thereof and having an average carbohydrate composition which differs from that of naturally-occurring human erythropoietin.

42. Vertebrate cells which can be propagated _in vitro_ continuously and which upon growth in culture are capable of producing in the medium of their growth in excess of 100 U of erythropoietin per 10^6 cells in 48 hours as determined by radioimmunoassay.

43. Vertebrate cells according to claim 42 capable of producing in excess of 500 U erythropoietin per 10^6 cells in 48 hours.

44. Vertebrate cells according to claim 42 capable of producing in excess of 1,000 U erythropoietin per 10^6 cells in 48 hours.

45. Vertebrate cells according to claim 42 which are mammalian or avian cells.

46. Vertebrate cells according to claim 45 which are COS-1 cells or CHO cells.

47. A synthetic polypeptide having part or all of the amino acid sequence as set forth in Table V and having one or more of the _in vivo_ or _in vitro_ biological activities of naturally-occurring monkey erythropoietin.

48. A synthetic polypeptide having part or all of the amino acid sequence set forth in Table VI, other than a sequence of residues entirely within the sequence numbered 1 through 20, and having a biological property of naturally-occurring human erythropoietin.

49. A synthetic polypeptide having part or all of the secondary conformation of part or all of the amino acid sequence set forth in Table VI, other than a sequence of residues entirely within the sequence numbered 1 through 20, and having a biological property of naturally-occurring human erythropoietin.

50. A process for the production of a polypeptide having part or all of the primary structural conformation and one or more of the biological properties of naturally-occurring erythropoietin, said process comprising:

growing, under suitable nutrient conditions, procaryotic or eucaryotic host cells transformed or transfected with a DNA vector according to claim 37, and isolating desired polypeptide products of the expression of DNA sequences in said vector.

51. An antibody substance characterized by immunoreactivity with erythropoietin and with a synthetic polypeptide having a primary structural conformation substantially duplicative of a continuous sequence of amino acid residues extant in naturally-occurring erythropoietin except for any polypeptide comprising a sequence of amino acid residues entirely comphrended within sequence,
A-P-P-R-L-I-C-D-S-R-V-L-E-R-Y-L-L-E-A-K.

52. An antibody according to claim 51, which is a monoclonal antibody.

53. An antibody according to claim 51, which is a polyclonal antibody.

54. An antibody according to claim 51, which is immunoreactive with erythropoietin and a synthetic polypeptide having the sequence selected from the sequences:
V-P-D-T-K-V-N-F-Y-A-W-K-R-M-E-V-G,
K-E-A-I-S-P-P-D-A-A-S-A-A, and
V-Y-S-N-F-L-R-G-K-L-K-L-Y-T-G-E-A-C-R-T-G-D-R.

55. A pharmaceutical composition comprising an effective amount of a polypeptide according to claims 1, 16, 39, 40 or 41 and a pharmaceutically acceptable diluent, adjuvant or carrier.

56. A method for providing erythropoietin therapy to a mammal comprising administering an effective amount of a polypeptide according to claims 1, 16, 39, 40 or 41.

57. A method according to claim 56 wherein the therapy comprises enhancing hematocrit levels.

58. A purified and isolated DNA sequence as set out in Table V or VI or a fragment thereof or the complementary strand of such a sequence or fragment.

59. A polypeptide product of the expression of a DNA sequence according to claim 58 in a procaryotic or eucaryotic host cell.

60. An improvement in the method for detection of a specific single stranded polynucleotide of unknown sequence in a heterogeneous cellular or viral sample including multiple single-stranded polynucleotides wherien:

(a) a mixture of labelled single-stranded polynucleotide probes is prepared having uniformly varying sequences of bases, each of said probes being potentially specifically complementary to a sequence of bases which is putatively unique to the polynucleotide to be detected,

(b) the sample is fixed to a solid substrate;

(c) the substrate having the sample fixed thereto is treated to diminish further binding of polynucleotides thereto except by way of hybridization to polynucleotides in said sample,

(d) the treated substrate having the sample fixed thereto is transitorily contacted with said mixture of labelled probes under conditions facilitative of hybridization only between totally complementary polynucleotides, and,

(e) the specific polynucleotide is detected by monitoring for the presence of a hybridization reaction between it and a totally complementary probe within said mixture of labelled probes, as evidenced by the presence of a higher density of labelled material on the substrate at the locus of the specific polynucleotide in comparison to a background density of labelled material resulting from non-specific binding of labelled probes to the substrate,

said improvement comprising using in excess of 32 mixed probes and performance of one or more of the following:

(1) employing a nylon-based paper as said solid substrate;

(2) treating with a protease in step (c);

(3) employing individual labelled probe concentrations of approximately 0.025 picomoles; and

(4) employing as one of the hybridization conditions in step (d) stringent temperatures approaching to with 4°C away from the lowest calculated Td of any of the probes employed.

0148605
Application number: 84308654.7

DECLARATION PURSUANT TO RULE 28, PARAGRAPH 4, OF THE EUROPEAN PATENT CONVENTION

The applicant has informed the European Patent Office that, until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, the availability of the micro-organism(s) identified below, referred to in paragraph 3 of Rule 28 of the European Patent Convention, shall be effected only by the issue of a sample to an expert.

IDENTIFICATION OF THE MICRO-ORGANISMS

Accession numbers of the deposits:

ATCC 39881
ATCC 39882
ATCC 39932
ATCC 39934
ATCC 39933

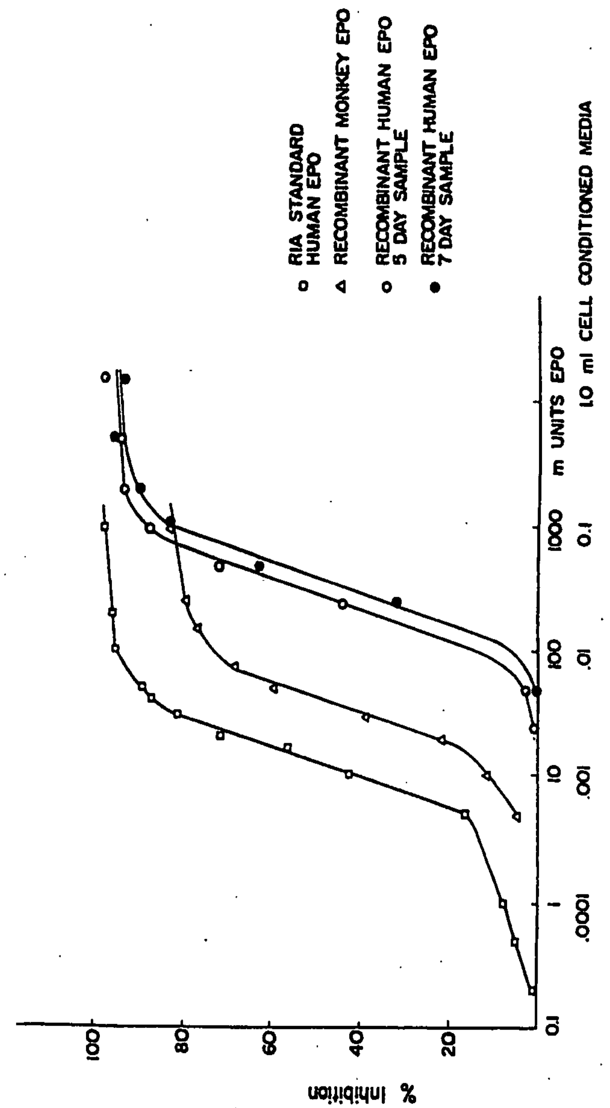

FIG. 1 Comparison of Recombinant Human & Monkey EPO in Radioimmunoassay

0148605

FIG.2

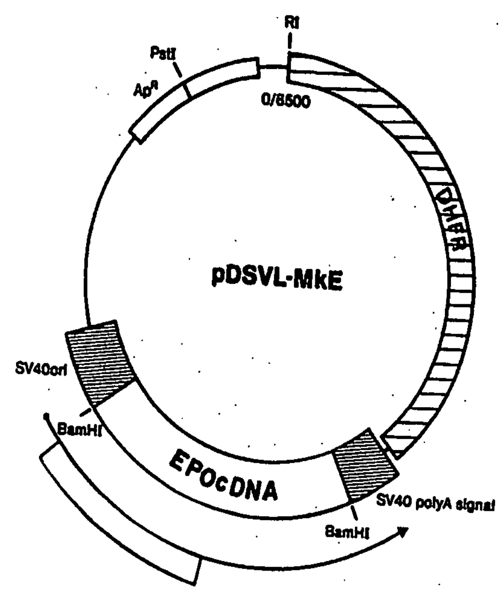

FIG.3

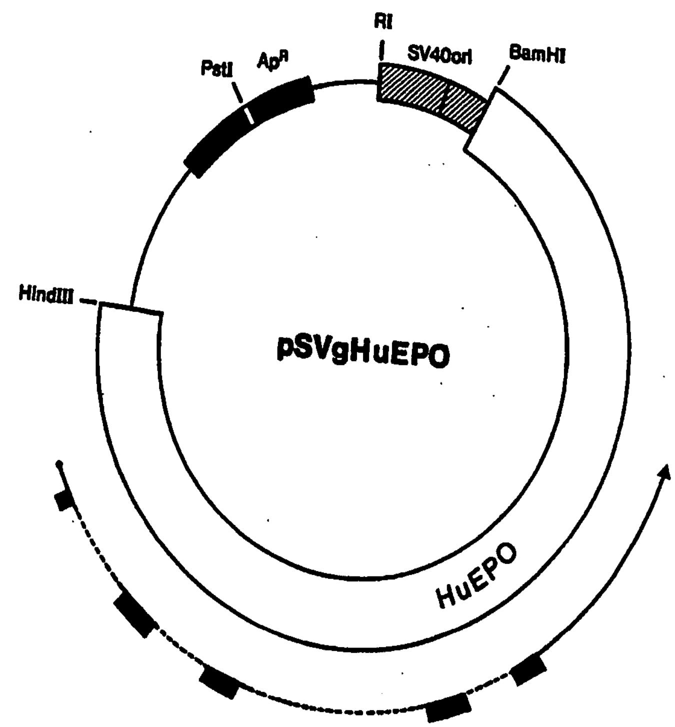

FIG. 4